04/00115

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18211 A1

(51) International Patent Classification[7]: C12N 15/29, C07K 14/415, A01H 5/00

(21) International Application Number: PCT/AU00/01033

(22) International Filing Date:
1 September 2000 (01.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PQ 2625 2 September 1999 (02.09.1999) AU

(71) Applicant *(for all designated States except US)*: THE UNIVERSITY OF QUEENSLAND [AU/AU]; St Lucia, Brisbane, QLD 4072 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: POTIER, Bernard [FR/ZA]; Institute For Plant Biotechnology, University Of Stellenbosch, Private Bag X1, 7602 Matieland (ZA). BIRCH, Robert, George [AU/AU]; 24 Yarrimbah Street, Jindalee, Brisbane, QLD 4074 (AU).

(74) Agent: FISHER ADAMS KELLY; Level 13, AMP Place, 10 Eagle Street, Brisbane, QLD 4000 (AU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/18211 A1

(54) Title: SUGARCANE PLANT PROMOTERS TO EXPRESS HETEROLOGOUS NUCLEIC ACIDS

(57) Abstract: Isolated nucleic acids corresponding to sugarcane plant gene promoters are provided. The promoters are capable of directing expression in many or all cells of a plant, preferentially in stem tissue or preferentially in meristem tissue. The isolated promoters are used in the production of transgenic plants, preferably transgenic sugarcane, to direct expression of heterologous nucleic acids in either a constitutive or tissue-specific manner.

TITLE

Sugarcane plant promoters to express heterologous nucleic acids.

FIELD OF THE INVENTION

THIS INVENTION relates generally to isolated nucleic acid promoters for use in plant genetic engineering. More particularly, the present invention relates to constitutive and tissue-specific promoters for expression of heterologous nucleic acids, such as foreign or endogenous coding sequences, in monocotyledonous plants. The invention is also concerned with a chimeric nucleic acid construct comprising a promoter of the invention operably linked to the heterologous nucleic acid, expression vectors, transformed plants, cells and tissues comprising an isolated promoter of the invention.

BACKGROUND OF THE INVENTION

A primary goal of genetic engineering is to obtain plants having improved characteristics or traits. Many different types of characteristics or traits are considered advantageous, but those of particular importance include resistance to plant diseases, resistance to viruses or insects and resistance to herbicides. Other advantageous characteristics or traits include tolerance to cold or soil salinity, enhanced stability or shelf life of the ultimate consumer product obtained from a plant, or improvement in the nutritional value of edible portions of a plant.

Recent advances in genetic engineering have enabled the incorporation of a selected gene (or genes) into plant cells to impart a desired quality (or qualities) to a plant of interest. The selected gene (or genes) may be derived from a source different from the plant of interest or may be native to the desired plant, but engineered to have different or improved qualities. This new gene (or genes) may then be expressed in cells of the regenerated plant to exhibit the new trait or characteristic.

In order for the newly incorporated gene to express the protein for which it codes in a plant cell, the proper regulatory signals must be present and in the proper location with respect to the gene. These regulatory signals include a promoter, a 5' non-translated leader sequence and a 3' polyadenylation signal.

Generally, the efficiency of gene expression is governed largely by the promoter used to express the gene. A promoter is a DNA sequence that directs the cellular machinery of a plant to produce (transcribe) RNA (transcript) from a contiguous transcribable region downstream (3') of the promoter. The promoter influences the rate at which the transcript of the gene is made. Assuming the transcript includes a coding region with appropriate translational signals, the promoter also influences the rate at which the resultant protein product of the gene is produced. Promoter activity also can depend on the presence of several other *cis*-acting regulatory elements which, in conjunction with cellular factors, determine strength, specificity, and transcription initiation site (for a review, see Zawel and Reinberg, 1992, *Curr. Opin. Cell Biol.* 4:488).

It has been shown that certain promoters are able to direct RNA synthesis at a higher rate relative to other promoters. These are called "strong promoters". Certain other promoters have been shown to direct RNA production at higher levels only in particular types of cells or tissues and are often referred to as "tissue-specific promoters". Promoters that are capable of directing RNA production in many or all tissues of a plant are called "constitutive promoters". Thus, expression of a chimeric gene (or genes) introduced into a plant may potentially be controlled by identifying and using a promoter with the desired characteristics.

A myriad of promoters has been described for gene expression in dicotyledonous plants. However, there is currently a dearth of promoters that can be used for effective expression of foreign or endogenous coding sequences in monocotyledonous plants.

SUMMARY OF THE INVENTION

Thus, in one aspect, the invention provides an isolated nucleic acid comprising a nucleotide sequence which corresponds to a promoter region of a transcribable DNA sequence that is hybridizable to a probe or primer derivable from a polynucleotide sequence selected from the group consisting of: -

(a) the polynucleotide sequence set forth in FIG. 14 [SEQ ID NO: 1];

(b) the polynucleotide sequence set forth in FIG. 15 under designator c51 [SEQ ID NO: 2];

(c) the polynucleotide sequence set forth in FIG. 15 under designator c511 [SEQ ID NO: 3];

(d) the polynucleotide sequence set forth in FIG. 15 under designator c512 [SEQ ID NO: 4];

(e) the polynucleotide sequence set forth in FIG. 16 [SEQ ID NO: 5];

(f) the polynucleotide sequence set forth in FIG. 17 [SEQ ID NO: 6];

(g) the polynucleotide sequence set forth in FIG. 23 [SEQ ID NO: 7];

(h) the polynucleotide sequence set forth in FIG. 24 [SEQ ID NO: 8];

(i) the polynucleotide sequence set forth in FIG. 26 [SEQ ID NO: 9];

(j) the polynucleotide sequence set forth in FIG. 29 [SEQ ID NO: 10]; and

(k) the polynucleotide sequence set forth in FIG. 31 [SEQ ID NO: 11].

Advantageously, the isolated promoter region of the transcribable DNA sequence is of sufficient length such that it is capable of initiating and regulating transcription of a DNA sequence to which it is coupled. The promoter region may be between 100 bp and 4 kb in length and preferably greater than 1 kb in length. An analogous promoter region can be obtained from any plant species that has a DNA sequence that is transcribed in one or more cells or tissues of the plant provided that the DNA sequence is capable of hybridising to a nucleic acid probe derivable from a polynucleotide sequence as set forth in any one of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

The polynucleotide sequences set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 are transcribable sequences obtained from sugarcane (*Saccharum* sp.). The polynucleotide sequences identified by SEQ ID NO: 1, 2, 3, 4, 9, 10 and 11 are highly transcribed in stem tissue, particularly mature stem tissue of sugarcane. Homologous sequences corresponding to SEQ ID NO: 2, 3 and 4 are expressed in

stem tissue of other monocotyledonous plants including maize and sorghum. Homologous sequences corresponding to SEQ ID NO: 1, 9, 10 and 11 are also suspected to exist in other monocotyledonous plants.

The invention thus also provides a polynucleotide sequence or variant thereof that is highly transcribed in stem tissue of monocotyledonous plants wherein said sequence is selected from the group consisting of any one of the polynucleotides set forth in SEQ ID NO: 1, 2, 3, 4, 9, 10 and 11.

The polynucleotide sequences identified by SEQ ID NO: 7 and 8 are highly transcribed in meristem tissue of sugarcane. Thus, the invention also provides a polynucleotide sequence or variant thereof that is highly transcribed in meristem tissue of monocotyledonous plants wherein said sequence is selected from the group consisting of SEQ ID NO: 7 or SEQ ID NO: 8.

The polynucleotides identified by SEQ ID NO: 5 and 6 are expressed constitutively in leaves, stems and roots of sugarcane. Homologous sequences corresponding to these polynucleotides are expressed in many or all tissues of other monocotyledonous plants including maize, rice and sorghum.

Accordingly, the invention also features a polynucleotide sequence or variant thereof that is highly transcribed constitutively in tissue of monocotyledonous plants wherein said sequence is selected from the group consisting of: -

(a) the polynucleotide sequence set forth in SEQ ID NO: 5; and

(b) the polynucleotide sequence set forth in SEQ ID NO: 6.

The foregoing polynucleotide sequences represent transcribable sequences, which can be used to construct a probe or primer(s) to isolate homologous transcribable sequences in other plant species, preferably, monocotyledonous species, so that a corresponding promoter region from said other plant species having the same tissue-specific or constitutive qualities can be isolated and used.

Preferably, the isolated nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) the polynucleotide sequence set forth in FIG. 18 [SEQ ID NO: 12];

(b) the polynucleotide sequence set forth in FIG. 19 [SEQ ID NO: 13];

(c) the polynucleotide sequence set forth in FIG. 20 [SEQ ID NO: 14];

(d) the polynucleotide sequence set forth in FIG. 21 [SEQ ID NO: 15];

(e) a biologically-active portion of any one of SEQ ID NO: 12, 13, 14 or 15; and

(f) a variant of any one of the polynucleotide sequences according to (a) to (e).

In one embodiment, the variant has at least 60%, preferably at least 80%, more preferably at least 90%, and most preferably at least 95% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 12, 13, 14 or 15.

In another embodiment, the variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 12, 13, 14 or 15 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In another aspect, the invention provides a chimeric gene comprising an isolated nucleic acid of the invention, variant or biologically-active fragment and a heterologous nucleic acid.

In yet another aspect, the invention provides an expression vector comprising an isolated nucleic acid of the invention, variant or biologically-active fragment.

Preferably, the expression vector comprises an isolated nucleic acid of the invention, variant or biologically-active fragment and a heterologous nucleic acid operably linked thereto.

In one embodiment, the heterologous nucleic acid encodes a polypeptide such as a structural or regulatory protein.

In another embodiment, the heterologous nucleic acid encodes a molecule which is capable of modulating expression of a target gene.

In a preferred embodiment, the molecule is an antisense RNA or a ribozyme or other transcribed region aimed at downregulation of expression of the corresponding target gene. For example, the said other transcribed region may

comprise a sense transcript aimed at sense suppression (co-suppression) of the target gene.

Depending on the polynucleotide selected, in one embodiment, the expression construct may be further characterized in that said isolated nucleic acid is capable of directing transcription of the heterologous nucleic acid preferentially in stem tissue of a plant. In another embodiment, expression construct may be further characterized in that that said isolated nucleic acid is capable of directing transcription of the heterologous nucleic acid preferentially in meristem tissue of a plant. In yet another embodiment, the expression construct may be further characterized in that said isolated nucleic acid is capable of directing transcription of the heterologous nucleic acid in many or all tissues of a plant.

In a further aspect, the invention provides a method of transforming a plant, including the step of introducing into a plant cell or tissue an isolated nucleic acid, chimeric gene or expression vector of the invention.

According to a still further another aspect of the invention, there is provided a transformed plant cell or tissue comprising an isolated nucleic acid, chimeric gene or expression vector of the invention.

In a still yet further aspect, the invention provides a transformed plant comprising an isolated nucleic acid, chimeric gene or expression vector of the invention.

Plants encompass any taxonomic grouping thereof, including angiosperms, gymnosperms, monocotyledons and dicotyledons. Preferred plants are monocotyledons such as cereals, sugarcane, bananas and pineapples, but without limitation thereto.

More preferably, the plant is sugarcane.

Preferably, the transformed plant has an altered phenotype compared to a corresponding non-transformed plant.

Preferably, the altered phenotype results from expression of the heterologous nucleic acid.

The invention also provides cells, tissues, leaves, fruit, flowers, seeds and other reproductive material, material used for vegetative propagation, progeny plants including F1 hybrids, male-sterile plants and all other plants and plant products derivable from transgenic plants of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the invention may be readily understood and put into practical effect, preferred embodiments will now be described by way of example with reference to the accompanying drawings in which:

FIG. 1: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, NCo310, Q110 and Q145. The probe is the cDNA c67 (SEQ ID NO: 1), revealing preferential expression of gene 67 in the stems of all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 0.7 /100 / 5 (Pindar), 0.3 / 100 / 1.7 (NCo310), 0.3 / 100 / 2.8 (Q110) and 0.7 / 100 / 2 (Q145).

FIG. 2: Northern analysis of total RNA extracted from different internodes of field-grown sugarcane plants (cultivar Pindar). Each plant had a total of 17 internodes, IN_1 corresponds to internode 1 (top of the stem) with other internodes numbered down the stem. The probe is the cDNA c67 (SEQ ID NO: 1), revealing preferential expression of gene 67 in mature internodes. The relative expression levels are 0 / 0.7 / 9.7 / 47 / 100 / 45 (IN_1 / IN_4 / IN_6 / IN_8 / IN_{11} / IN_{15}).

FIG. 3: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, NCo310, Q110 and Q145. The probe is a *Nhe*I DNA fragment from the cDNA c51 (SEQ ID NO: 2, 3 and 4), revealing stem-preferential expression of gene 51 in all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 19.5 / 100 / 18 (Pindar), 17.5 / 100 / 23.7 (NCo310), 16.8 / 100 / 20 (Q110) and 12.5 / 100 / 15.8 (Q145).

FIG. 4: Northern analysis of total RNA extracted from different parts of cultivar Pindar sugarcane plants with 17 internodes. IN_6 and IN_{11} correspond respectively to internode 6 and 11. Nc is the core of a nodal part of the stem and N is

an entire nodal part of the stem. M corresponds to the meristematic region of the stem. R is the root tissue. YL, ML and OL represent different stages of leaf development, Y being the young and recently emerged leaves, M being the more mature leaves and O being leaves at an older stage. The probe is the cDNA c51 (SEQ ID NO: 2, 3 and 4), revealing expression of gene 51 throughout the sugarcane stem.

FIG. 5: Northern analysis of total RNA extracted from different parts of field-grown maize plants and glasshouse-grown sorghum plants. The different tissues analysed were male flower parts (F), top of the stem (TS), medium and lower part of the stem (MS and LS), root system (R), leaves at a young, mature and older stage of development (YL, ML and OL) and panicle (P). The probe is the cDNA c51, revealing expression of a homologue of sugarcane stem-specific gene 51 (SEQ ID NO: 2, 3 and 4) in maize and sorghum.

FIG. 6: Northern analyses of total RNA extracted from different parts of field-grown sugarcane (cultivar Pindar) and maize (sweet corn) stems. For the sugarcane RNA, each plant had a total of 17 internodes, IN_1 corresponds to internode 1 (top of the stem), with other internodes numbered down the stem. For the maize RNA, each plant had a total of 13 internodes, IN_1 corresponds to internode 1 (top of the stem), with other internodes numbered down the stem. The probe is the cDNA c51 (SEQ ID NO: 2, 3 and 4), revealing expression of gene 51 in internodes at all stages of maturity in both sugarcane and maize. The relative expression levels are 44 / 47 / 51 / 64 / 100 / 61 (IN_1 / IN_4 / IN_6 / IN_8 / IN_{11} / IN_{15}) in sugarcane stems and 40 / 44 / 75 / 87 / 100 / 72 (IN_1 / IN_3 / IN_5 / IN_7 / IN_9 / IN_{11}) in maize stems.

FIG. 7: Northern analysis of total RNA extracted from various parts of field-grown sugarcane plants (cultivar Pindar). Each plant had a total of 17 internodes. The tissues analysed were a nodal part of the stem (N), internodes 1-2 (IN_{1-2}, top of the stem), 6 and 11 (IN_6 and IN_{11}), the root system (R) and leaves at a young, mature and older stage of development (YL, ML and OL). The probe is the cDNA c32A (SEQ ID NO: 5 and 6), revealing expression of gene 32A in all tested sugarcane tissues. The relative expression levels are 42 / 100 / 92 / 94 / 59 / 43 / 70 / 60 (N / IN_{1-2} / IN_6 / IN_{11} / R / YL / ML / OL).

FIG. 8: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, NCo310, Q110 and Q145. The probe is the cDNA c322 (SEQ ID NO: 5 and 6), revealing constitutive expression of gene 32A in all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 58 / 100 / 40 (Pindar), 48 / 100 / 43 (NCo310), 88 / 100 / 58 (Q110) and 70 / 100 / 51 (Q145).

FIG. 9: Northern analyses of total RNA extracted from different parts of field-grown sugarcane (cultivar Pindar) and maize (sweet corn) stems. For the sugarcane RNA, each plant had a total of 17 internodes, IN_1 corresponds to internode 1 (top of the stem), with other internodes numbered down the stem. For the maize RNA, each plant had a total of 13 internodes, IN_1 corresponds to internode 1 (top of the stem), with other internodes numbered down the stem. The probe is the cDNA c322 (SEQ ID NO: 5 and 6), revealing expression of gene 32A and its maize homologue in internodes at all stages of maturity in both sugarcane and maize. The relative expression levels are 81 / 71 / 81 / 98 / 100 / 74.5 (IN_1 / IN_4 / IN_6 / IN_8 / IN_{11} / IN_{15}) in sugarcane stems and 74.5 / 94 / 97.5 / 100 / 82 / 90 (IN_1 / IN_3 / IN_5 / IN_7 / IN_9 / IN_{11}) in maize stems.

FIG. 10: Northern analysis of total RNA extracted from glasshouse-grown sorghum and rice plants. For the sorghum RNA, the different tissues analysed were the panicle (P), the top, middle part and lower parts of the stem (TS, MS and LS), the root system (R) and leaves at a young, mature and older stage of development (YL, ML and OL). For the rice RNA, the different tissues analysed were the young and older stems (S_2 and S_1), the root system (R) and the young, mature and older leaves (L_3, L_2 and L_1). The probe is the cDNA c322 (SEQ ID NO: 5 and 6), revealing a constitutive expression of homologues of gene 32A in all the tissues tested in sorghum and rice.

FIG. 11: Southern analysis of genomic DNA (10 µg/lane) from cultivar Q117 digested by *Eco*RI (E), *Kpn*I (K), *Sac*I (S), *Xba*I (X) and *Not*I (N), indicating genomic copies of gene 67 in sugarcane. The probe is the cDNA c67 (SEQ ID NO: 1). Lane λ_{H3} indicates the position of molecular weight markers on the same electrophoresis gel.

FIG. 12: Southern analysis of genomic DNA (10 µg/lane) from cultivar Q117 digested by *Eco*RI (E), *Kpn*I (K), *Sac*I (S), *Xba*I (X) and *Not*I (N), indicating few genomic copies of gene 51 in sugarcane. The probe is the 3' region of the cDNA c51 (nucleotide 621 to nucleotide 874 of SEQ ID NO: 2, 3 and 4). Lane λ_{H3} indicates the position of molecular weight markers on the same electrophoresis gel.

FIG. 13: Southern analysis of genomic DNA (10 µg/lane) from cultivar Q117 digested by *Eco*RI (E), *Kpn*I (K), *Sac*I (S), *Xba*I (X) and *Not*I (N), indicating multiple copies of gene 32A in sugarcane. The probe is the cDNA c322 (SEQ ID NO: 5 and 6). Lane λ_{H3} indicates the position of molecular weight markers on the same electrophoresis gel.

FIG. 14: Nucleotide sequence of the cDNA c67 (SEQ ID NO: 1) and deduced amino acid sequence of the longest open reading frame (SEQ ID NO: 16). The ATG start codon is at position 33 to 35, the stop codon (TGA) is at position 594 to 596. No homologous sequences (with more than 40% identity on the whole sequence) were found in the databases (NR nucleic, ANGIS) at either the nucleotide or the amino acid level. Two putative polyadenylation signals (AATAAA) are present between nucleotide 853 and nucleotide 862.

FIG. 15: Nucleotide sequence alignment of the cDNA clones c51, c511 and c512 (SEQ ID NO: 2, 3 and 4) and deduced amino acid sequence (aa) of the longest and conserved open reading frames (SEQ ID NO: 17). The ATG start codon has been assigned to position 129 to 131, the stop codon (TGA) is at position 576 to 578. The cDNA clone c511 has a stop codon TAG at position 435 to 437. The alanine (A) residue deduced from nucleotide position 402-404 in the cDNAs c51 and c511 is changed to a valine residue (V) in the deduced amino acid sequence from the cDNA c512. The valine (V) residue deduced from nucleotide position 501-503 in the cDNAs c511 and c512 is changed to an alanine (A) residue in the deduced amino acid sequence of the cDNA c51. The nucleotide and deduced amino-acid sequences of all the cDNAs have no homologues in the nucleotide and protein databases (NR nucleic, ANGIS).

FIG. 16: Nucleotide sequence of cDNA clone c32A (SEQ ID NO: 5) and deduced amino acid sequence (SEQ ID NO: 22). The stop codon (TAG) is at position 44 to 46. This partial protein sequence (SEQ ID NO: 22) is homologous to S27 ribosomal proteins from rice, barley, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii* and other organisms.

FIG. 17: Nucleotide sequence and deduced amino acid sequence of the cDNA clone c322 (SEQ ID NO: 6 and 23, respectively). The ATG start codon is at position 92 to 94, the stop codon (TAG) is at position 350 to 352. The protein sequence (SEQ ID NO: 23) is homologous to S27 ribosomal proteins from rice, barley, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii* and other organisms.

FIG. 18: Nucleotide sequence of the 67pro promoter sequence (SEQ ID NO: 12). The start of the coding sequence is at position 1045. A putative "tata box" is present at position 964-969. The predicted start of transcription is at position 994 (A).

FIG. 19: Nucleotide sequence of the 32a2pro promoter sequence (SEQ ID NO: 13). The start of the coding sequence is at position 1943.

FIG. 20: Nucleotide sequence of the 32a6pro promoter sequence (SEQ ID NO: 14). The start of the coding sequence is at position 1266. A putative "tata box" is present at position 1138-1145. The predicted start of transcription is at position 1170 (C).

FIG. 21: Nucleotide sequence of the 51pro promoter sequence (SEQ ID NO: 15). The start of the coding sequence is at position 2946.

FIG. 22: GUS histochemical assay of cross-sections of three transgenic sugarcane plant stems transformed with plasmid p67G. The strong blue staining (shown as dark grey to black on the figure) indicates the presence of the GUS protein and confirms the functionality of the promoter fused to the GUS gene.

FIG. 23: Nucleotide sequence of the cDNA c19 (SEQ ID NO: 7). Neither the nucleotide sequence nor the deduced amino acid sequences (in the six frames) have homologues in the databases (NR nucleic and NR Proteins, ANGIS).

FIG. 24: Nucleotide sequence of the cDNA c3 (SEQ ID NO: 8). The sequence from nucleotide 144 to nucleotide 395 (deduced amino acid sequence SEQ ID NO: 21) is homologous to part of the cellulase (EC 3.2.1.4) coding sequence from a variety of sources (*e.g. Arabidopsis thaliana*, tomato, tobacco, pepper, termites and bacteria).

FIG. 25: Northern analysis of total RNA extracted from different parts of field-grown sugarcane plants. The probe is the cDNA c19 (SEQ ID NO: 7), revealing preferential expression in the meristematic region of the stem (M). No expression can be detected in other tissues such as the core region of a stem internode (INc), a whole stem internode (IN), the core region of a stem node (Nc), a whole stem node (N), the root system (R), young leaves (YL), mature leaves (ML) and older leaves (OL). An identical result is obtained (except for the size of the hybridising transcript) when using the cDNA c3 (SEQ ID NO: 8) as the probe.

FIG. 26: Nucleotide sequence and deduced amino acid sequence of the cDNA c18 (SEQ ID NO: 9 and 18, respectively). The ATG start codon is present at position 1-3 and the stop codon (TGA) is at position 244-246. The amino acid sequence is homologous to type 2 plant metallothionein-like proteins previously described in different plants such as *Arabidopsis thaliana*, tobacco, barley, fava bean and others.

FIG. 27: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, NCo310, Q110 and Q145. The probe is the cDNA c18 (SEQ ID NO: 9), revealing preferential expression of gene 18 in the stems of all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 6 / 100 / 8.5 (Pindar), 9 / 100 / 10 (NCo310), 11.5 / 100 / 11.5 (Q110) and 6 / 100 / 5 (Q145).

FIG. 28: Northern analysis of total RNA extracted from different internodes of field-grown sugarcane plants (cultivar Pindar). Each plant had a total of 17 internodes, IN_1 corresponds to internode 1 (top of the stem) with other internodes numbered down the stem. The probe is the cDNA c18 (SEQ ID NO: 9), revealing preferential expression in the young parts of the stem, with a gradual decrease in the expression levels towards the base of the stem (older parts). The

relative expression levels are 100 / 98 / 86 / 69 / 40 / 25 (IN_1 / IN_4 / IN_6 / IN_8 / IN_{11} / IN_{15}).

FIG. 29: Nucleotide sequence of the cDNA c53A (SEQ ID NO: 10) and deduced amino acid sequence (SEQ ID NO: 19). The ATG start codon is at position 124-126, the stop codon TGA is at position 1069-1071. The protein sequence is homologous to homeobox proteins described in plants (*e.g. Arabidopsis thaliana* ATK1 protein and KNA2 knotted-like homeobox protein, soybean HMB1 homeobox protein, maize RS1 gene product, rice OSH1 homeobox 1 protein and others).

FIG. 30: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, Nco310, Q110 and Q145. The probe is the cDNA c53A (SEQ ID NO: 10), revealing preferential expression of gene 53A in the stems of all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 4 / 100 / 14 (Pindar), 5 / 100 / 12 (NCo310), 0 / 100 / 11 (Q110) and 0 / 100 / 8 (Q145).

FIG. 31: Nucleotide sequence (SEQ ID NO: 11) and deduced amino acid sequence (SEQ ID NO: 20) of the cDNA c57. The stop codon (TGA) is at position 567-569. The amino acid sequence and the nucleotide coding sequence are homologous to several auxin-induced and auxin-responsive genes / proteins from different plants such as *Arabidopsis thaliana*, soybean and tobacco.

FIG. 32: Northern analysis of total RNA extracted from leaf (L), stem (S) and root (R) of field-grown sugarcane plants. The cultivars were Pindar, Nco310, Q110 and Q145. The probe is the cDNA c57 (SEQ ID NO: 11), revealing preferential expression of gene 57 in the stems of all tested sugarcane cultivars. The relative expression levels for the leaf / stem / root tissues are 0 / 100 / 0 (Pindar), 0 / 100 / 0 (Nco310), 15 / 100 / 7 (Q110) and 15 / 100 / 12 (Q145).

FIG. 33: Northern analysis of total RNA extracted from different internodes of field-grown sugarcane plants (cultivar Pindar). Each plant had a total of 17 internodes, IN_1 corresponds to internode 1 (top of the stem) with other internodes numbered down the stem. The probe is the cDNA c57 (SEQ ID NO: 11), revealing preferential expression in the young parts of the stem, with a gradual

decrease in the expression levels towards the base of the stem (older parts). The relative expression levels are 100 / 86 / 82 / 65 / 45 / 20 (IN_1 / IN_4 / IN_6 / IN_8 / IN_{11} / IN_{15}).

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 Nucleotide sequence of cDNA clone c67, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 2 Nucleotide sequence of cDNA clone c51, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 3 Nucleotide sequence of cDNA clone c511, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 4 Nucleotide sequence of cDNA clone c512, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 5 Nucleotide sequence of cDNA clone c32A, which is transcribed constitutively in leaves, stems and roots of field-grown sugarcane plants.

SEQ ID NO: 6 Nucleotide sequence of cDNA clone c322, which is transcribed constitutively in leaves, stems and roots of field-grown sugarcane plants.

SEQ ID NO: 7 Nucleotide sequence of cDNA clone c19, which is preferentially expressed in meristematic tissue of field-grown sugarcane plants.

SEQ ID NO: 8 Nucleotide sequence of the cDNA c3, which is preferentially expressed in meristematic tissue of field-grown sugarcane plants.

SEQ ID NO: 9 Nucleotide sequence of cDNA clone c18, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 10 Nucleotide sequence of cDNA clone c53A, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 11 Nucleotide sequence of cDNA clone c57, which is preferentially expressed in the stem of field-grown sugarcane plants.

SEQ ID NO: 12 Nucleotide sequence of the stem-specific 67 promoter. *E. coli* DH10B strain transformed with plasmid pG4-67pro comprising the 67 promoter was deposited with AGAL on August 30, 1999 under accession number NM99/05995.

SEQ ID NO: 13 Nucleotide sequence of the stem-specific 32A2 promoter. *E. coli* TOP10 strain transformed with pZ-32A2P2 comprising the 32A2 promoter was deposited with AGAL on August 30, 1999 under accession number NM99/05994.

SEQ ID NO: 14 Nucleotide sequence of the constitutive 32A6 promoter. *E. coli* TOP10 strain transformed with pZ-32A6P2 comprising the 32A6 promoter was deposited with AGAL on August 30, 1999 under accession number NM99/05993.

SEQ ID NO: 15 Nucleotide sequence of the constitutive 51 promoter. *E. coli* TOP10 strain transformed with pZ-H51-6P1P2 comprising the 51 promoter was deposited with AGAL on August 30, 1999 under accession number NM99/05992.

SEQ ID NO: 16 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 1.

SEQ ID NO: 17 Deduced amino acid sequence of the longest and conserved open reading frame common to each of SEQ ID NO: 2, 3 and 4.

SEQ ID NO: 18 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 5.

SEQ ID NO: 19 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 10.

SEQ ID NO: 20 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 11.

SEQ ID NO: 21 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 8.

SEQ ID NO: 22 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 5.

SEQ ID NO: 23 Deduced amino acid sequence of the longest open reading frame of SEQ ID NO: 6.

DETAILED DESCRIPTION

1. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

"Amplification product" refers to a nucleic acid product generated by nucleic acid amplification techniques.

By *"biologically active fragment"* is meant a segment, portion or fragment of a the biological active molecule which has at least about 0.1%, preferably at least about 10%, more preferably at least about 25% and even more preferably at least 50% of the activity of the molecule.

"Chimeric gene" is defined herein as a nucleic acid, preferably a DNA molecule, either single- or double-stranded, which includes an isolated nucleic acid of the invention, variant or biologically-active fragment, operably linked to a heterologous nucleic acid.

By "*corresponds to*" or "*corresponding to*" is meant similar, related or analgous to in structure and/or function. For example, an isolated nucleic acid of the invention suitably comprises a nucleotide sequence structurally and/or functionally related to a corresponding promoter sequence, such that the isolated nucleic acid of the invention is capable of functioning as a promoter.

The term *"regeneration"* as used herein means growing a whole, differentiated plant from a plant cell, a group of plant cells, a plant part (including seeds), or a plant piece (e.g., from a protoplast, callus, or tissue part).

By *"heterologous nucleic acid"* is meant a nucleic acid distinct from an isolated promoter of the invention. Operationally, the heterologous nucleic acid is operably linked to an isolated promoter nucleic acid of the invention to achieve expression of the heterologous nucleic acid. The term heterologous nucleic acid encompasses transcribable DNA as will be defined hereinafter.

"Homology" refers to the percentage number of nucleotides of a polynucleotide sequence that are identical to a reference polynucleotide sequence. Homology may be determined using sequence comparison programs such as BESTFIT (Deveraux *et al.* 1984, *Nucleic Acids Research* 12, 387-395) which is incorporated herein by reference. In this way sequences of a similar or substantially different length to those cited herein might be compared by insertion of gaps into the alignment, such gaps being determined, for example, by the comparison algorithm used by BESTFIT.

"Hybridization" is used herein to denote the pairing of complementary nucleotide sequences to produce a DNA-DNA hybrid or a DNA-RNA hybrid. Complementary base sequences are those sequences that are related by the base-pairing rules. In DNA, A pairs with T and C pairs with G. In RNA U pairs with A and C pairs with G. In this regard, the terms "match" and "mismatch" as used herein refer to the hybridisation potential of paired nucleotides in complementary nucleic acid strands. Matched nucleotides hybridise efficiently, such as the classical A-T and G-C base pair mentioned above. Mismatches are other combinations of nucleotides that do not hybridise efficiently.

By *"isolated"* is meant material that is substantially or essentially free from components that normally accompany it in its native state. For example, an "isolated nucleic acid", as used herein, refers to a nucleic acid, which has been purified from the sequences which flank it in a naturally occurring state, *e.g.*, a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment.

By *"marker gene"* is meant a gene that imparts a distinct phenotype to cells expressing the marker gene and thus allows such transformed cells to be distinguished from cells that do not have the marker. A selectable marker gene confers a trait for which one can 'select' based on resistance to a selective agent (*e.g.*, a herbicide, antibiotic, radiation, heat, or other treatment damaging to untransformed cells). A screenable marker gene (or reporter gene) confers a trait that one can identify through observation or testing, *i.e.*, by 'screening' (*e.g.* green fluorescent protein or enzymes such as β-glucuronidase, neomycin phosphotransferase II, and luciferase not present in untransformed cells).

The term *"nucleic acid"* as used herein designates DNA and/or RNA including mRNA, cRNA, cDNA and genomic DNA in single-stranded and double-stranded form, and encompasses oligonucleotides and polynucleotides as herein defined.

By *"obtained from"* is meant that a sample such as, for example, a nucleic acid extract is isolated from, or derived from, a particular source of the host. For example, the nucleic acid extract may be obtained from tissue isolated directly from the host.

The term *"oligonucleotide"* as used herein refers to a polymer composed of a multiplicity of nucleotide units (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotides and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule may vary depending on the particular application. An oligonucleotide typically comprises from about 10 to 30 nucleotides, but the term can refer to molecules of any length, although the term *"polynucleotide"* usually refers to nucleic acids of more than 100 nucleotides in length.

By *"operably linked"* is meant functionally linked. For example, an

isolated nucleic acid of the invention is operably linked to a heterologous nucleic acid regulated by being linked thereto so as to be capable of initiating, controlling, regulating or otherwise directing transcription of the heterologous nucleic acid. Usually, the isol ated nucleic acid is located upstream or 5' of the heterologous nucleic acid.

"*Polypeptide*", "*peptide*" and "*protein*" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

By "*primer*" is meant an oligonucleotide or polynucleotide which is capable of hybridizing to a complementary or partly complementary nucleic acid strand (template) and initiating synthesis of a primer extension product in the presence of a suitable polymerising agent. The primer is preferably single-stranded for maximum efficiency in amplification but may alternatively be double-stranded. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerisation agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotides, although it may contain fewer nucleotides. Primers can be large polynucleotides, such as from about 200 nucleotides to several kilobases or more. Primers may be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridise and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridise with the template. Preferably, the primer contains no mismatches with the template to which it is designed to hybridise but this is not essential. For example, non-complementary nucleotides may be attached to the 5'-end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotides or a stretch of non-complementary nucleotides can be interspersed into a primer,

provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridise therewith and thereby form a template for synthesis of the extension product of the primer.

"Probe" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another nucleic acid, often called the "target nucleic acid", through complementary base pairing. Probes may bind target nucleic acids lacking complete sequence complementarity with the probe, depending on the stringency of the hybridisation conditions. Probes can be directly or indirectly labelled.

The term *"promoter"* refers to a nucleic acid which directs expression of another nucleic acid to which it is operably linked, by initiating, regulating or otherwise controling transcription of said another nucleic acid.

"Constitutive promoter" refers to a promoter that directs transcription in many or all tissues of a plant.

By *"stem-specific promoter"* is meant a promoter that preferentially directs transcription in stem tissue of a plant.

By *"meristem-specific promoter"* is meant a promoter that preferentially directs transcription in meristem tissue of a plant.

The term *"recombinant"* as used herein refers to a nucleic acid or polypeptide resulting from *in vitro* manipulation into a form not normally found in nature. As used in the art, *"recombinant"* usually refers to the products of recombinant DNA technology.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity" and "substantial identity". A *"reference sequence"* is at least 6 but frequently 15 to 18 and often at least 25 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (*i.e.*, only a portion of the complete polynucleotide sequence) that is similar between the two

polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "*comparison window*" refers to a conceptual segment of typically 6 to 12 contiguous residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerised implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA, incorporated herein by reference) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389, which is incorporated herein by reference. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons Inc, 1994-1998, Chapter 15.

The term *"sequence identity"* as used herein refers to the extent that sequences are identical on a nucleotide-by-nucleotide basis over a window of comparison. Thus, a *"percentage of sequence identity"* is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, *"sequence identity"* will be understood to mean the "match percentage" calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using

standard defaults as used in the reference manual accompanying the software, which is incorporated herein by reference.

"Stringency" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridisation. The higher the stringency, the higher will be the degree of complementarity between immobilised nucleotide sequences and the labelled polynucleotide sequence that remain bound following hybridization.

"Stringent conditions" refers to temperature and ionic conditions under which only nucleotide sequences having a high frequency of complementary bases will hybridise. The stringency required is nucleotide sequence dependent and also depends upon the various components present during hybridisation. Generally, stringent conditions are selected to be about 10 to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a target sequence hybridises to a complementary probe.

The term *"transcribable DNA sequence"* or *"transcribed DNA sequence"*, excludes the non-transcribed regulatory sequence that drives transcription. Depending on the aspect of the invention, the *transcribable* sequence may be derived in whole or in part from any source known to the art, including a plant, a fungus, an animal, a bacterial genome or episome, eukaryotic, nuclear or plasmid DNA, cDNA, viral DNA or chemically synthesized DNA. A *transcribable* sequence may contain one or more modifications in either the coding or the untranslated regions which could affect the biological activity or the chemical structure of the expression product, the rate of expression or the manner of expression control. Such modifications include, but are not limited to, insertions, deletions and substitutions of one or more nucleotides. The *transcribable* sequence may contain an uninterrupted coding sequence or it may include one or more introns, bound by the appropriate splice junctions. The *transcribable* sequence may also encode a fusion protein. It is contemplated that introduction into plant tissue of chimeric nucleic acid constructs of the invention will include constructions wherein the *transcribable* sequence and its promoter are each derived from different species.

The term *"transformation"* means alteration of genotype by introduction of genetic material into an organism.

By *"transgenic"* is meant an organism that is transformed.

By *"transgenote"* is meant an immediate product of a transformation process.

The term *"variant"* is used in the context of a nucleic acid or protein which displays a definable level of sequence identity with a reference nucleic acid or protein respectively. For example, a variant nucleic acid is hybridizable with a reference sequence under stringent conditions that are defined hereinafter, or shares a percent level of sequence identity definable using a sequence comparison algorithm as hereinbefore described. Variants also encompass nucleic acids in which one or more nucleotides have been added or deleted, or replaced with different nucleotides or modified bases (eg. inosine, methylcytosine). In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference nucleic acid whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide. The term *"variant"* also include naturally occurring allelic variants.

By *"vector"* is meant a nucleic acid, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a nucleic acid sequence may be inserted or cloned. A vector preferably contains one or more unique restriction sites and may be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integratable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system may comprise a single vector or plasmid, two or more vectors or

plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector may also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are well known to those of skill in the art.

Throughout this specification, unless the context requires otherwise, the words "comprise", "comprises" and "comprising" will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

2. Transcribed DNA sequences

The promoter regions of the present invention were discovered by their location adjacent to the start of DNA sequences that were found to be transcribed at high levels in one or more tissues of sugarcane (*Saccharum* sp.). A hybridisation screening procedure, as described hereinafter, was used to identify genes expressed differentially in various tissues. However, it will be understood that the present invention is not restricted to use of any particular method for identifying such differentially expressed genes. For example, alternative procedures for identifying genes expressed differentially in various tissues include, but are not restricted to: cDNA and genomic subtractive hybridisation as for example described by Bulman and Neill (1996, In *"Plant Gene Isolation: Principles and Practice"*, G.D. Foster and D. Twell, eds Chichester, UK, Wiley, pp 369-397); multi-probe fluorescent analysis of microscopic cDNA arrays as for example described by Schena (1996 *BioEssays* 18:427-431); mRNA differential display as for example described by Liang and Pardee (1992, *Science* 257:967-970) and by Callard *et al.* (1994, *BioTechniques* 16:1096-1103); computer analysis of mRNA abundance based on frequency of occurrence of identical sequences emerging from large-scale sequencing of cDNA ends (ESTs) as for example taught by Cooke *et al* (1996, *EST and genomic sequencing projects*. In *Plant Gene Isolation: Principles and Practice*, *supra*, pp. 410-419); or promoter tagging by insertional mutagenesis with promoterless reporter genes as for example disclosed by Lindsey and Topping (1996,

T-DNA-mediated insertional mutagenesis. In *Plant Gene Isolation: Principles and Practice, supra*, pp. 275-300) and Mudge and Birch (1998, *Austral. J. Plant Physiol.* 25:637-643), which are all incorporated herein by reference.

2.1. Stem-specific transcribed sequences

The invention further provides DNA sequences that are transcribed preferentially at high levels in stem tissue, as compared to other tissues, of sugarcane plants. Exemplary sequences of this type are set forth in SEQ ID NO: 1, 2, 3 and 4.

SEQ ID NO: 1 is a cDNA of 976 nucleotides in length. This sequence has a 5' untranslated region (5' UTR) situated at position 1 to 32 and a 3' untranslated region (3' UTR) at position 594 to 881. A poly(A) tail is present at the end of this sequence. Two putative polyadenylation signals (AATAAA) are present between nucleotide 853 and nucleotide 862. An open reading frame (ORF) is present from nucleotide 33 to nucleotide 593. The nucleotide sequence and the deduced amino acid sequence (FIG. 14, SEQ ID NO: 16) do not show any homologues in the databases with more than 40% overall sequence identity.

SEQ ID NO: 2, 3 and 4 are homologous cDNA sequences (FIG. 15). The 5' UTRs are present from nucleotide position 1 to 128 (FIG. 15) and the 3' UTRs from nucleotide position 576 to the end of the sequences. The three cDNA sequences have poly(A) tails and a putative polyadenylation signal (AATAAA) is present at nucleotide position 810 to 815 (FIG. 15). The longest ORF is present from nucleotide position 129 to 575 (FIG. 15). The deduced amino acid sequence (SEQ ID NO: 17) is also presented in FIG. 15. A mutation in SEQ ID NO: 3 at position 435 (FIG. 15) introduced an in-frame stop codon, but the presence of the SEQ ID NO: 3 in the cDNA library indicates that it is still being transcribed from a functional promoter. Screening of the databases (nucleotide sequences and protein sequences) only revealed one similar sequence: EST Zm474 (accession number W49474) from maize. This maize EST is similar (83% sequence identity, 3 gaps) to the nucleotide sequence between nucleotide position 692 and 887 (FIG. 15), which is located within the 3' UTR.

SEQ ID NO: 9 is a partial cDNA of 597 nucleotides in length. An ORF is present from nucleotide positions 1 to 243 (FIG. 26). Its deduced amino acid sequence (SEQ ID NO: 18) is homologous to the type 2 plant metallothionein-like proteins. Examples of such proteins can be found in the nucleic acid and protein databases under accession numbers P30564 (castor bean), P43391 and P25860 (*Arabidopsis thaliana*) and others. A 3' UTR is present from nucleotide position 244 to the end of the sequence, which includes a poly(A), tail (FIG. 26).

SEQ ID NO: 10 is a cDNA of 1339 nucleotides in length. An ORF is present from nucleotide positions 124 to 1068 (FIG. 29, SEQ ID NO: 19). A 5' UTR is present from nucleotide positions 1 to 123 and a 3' UTR is present from nucleotide positions 1069 to the end of the sequence (FIG. 29). A putative polyadenylation signal is present between nucleotide positions 1301 and 1306 (TATAAA, FIG. 29). The amino acid sequence deduced from the ORF (SEQ ID NO: 19) is homologous to homeobox genes / proteins that can be found in the databases (nucleic acids and proteins) under such accession numbers as AAD13611 (maize), G4887610 (rice), BAA76750 (tobacco) and others.

SEQ ID NO: 11 is a cDNA of 1046 nucleotides in length. This sequence has a partial ORF from nucleotide positions 1 to 566 (FIG. 31). Its deduced amino acid sequence (SEQ ID NO: 20) is homologous to auxin-induced or auxin-responsive proteins found in the databases (nucleic acids and proteins) under such accession numbers as, for example, AT18409 and AT18416 (*Arabidopsis thaliana*), GMAUX28 (soybean), AF123509 and AF123508 (tobacco) and others. A 3'UTR is present from nucleotide position 567 to the end of the sequence (FIG. 31), and includes multiple putative polyadenylation signals such as AAAAAA and AATAAT, and a poly(A) tail.

SEQ ID NO: 16 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 1 (FIG. 14) (see SEQ ID NO: 1).

SEQ ID NO: 17 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 2, 3 and 4 (FIG. 15) (see SEQ ID NO: 2, 3 and 4).

SEQ ID NO: 18 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 9 (FIG. 26) (see SEQ ID NO: 9).

SEQ ID NO: 19 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 10 (FIG. 29) (see SEQ ID NO: 10).

SEQ ID NO: 20 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 11 (FIG. 31) (see SEQ ID NO: 11).

2.2. Meristem-specific transcribed sequences

SEQ ID NO: 7 is a partial cDNA of 366 nucleotides in length (FIG. 23). No homologues could be found in the nucleic acids databases.

SEQ ID NO: 8 is a partial sequence of a cDNA. This sequence of 395 nucleotides is homologous to cellulase (EC 3.2.1.4) nucleotide sequences such as the ones found in the nucleic acids databases under the accession numbers, for example, THFE4AA and CFICENBAA (bacteria), AF128404 (tobacco) or SLU20590 (tomato). Part of this nucleotide sequence (from nucleotide 144 to nucleotide 395, FIG. 24) can be translated into an amino acid sequence homologous to cellulase amino acid sequence (FIG. 24, SEQ ID NO: 21)

SEQ ID NO: 16 is the amino acid sequence derived from an ORF in the nucleotide sequence of the SEQ ID NO: 1 (FIG. 14). This amino acid sequence has no homologues in the protein databases (see SEQ ID NO: 1).

SEQ ID NO: 21 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 8 (FIG. 24) (see SEQ ID NO: 8).

2.3. Constitutive transcribed sequences

The invention also features DNA sequences that are transcribed constitutively in leaves, stems and roots of sugarcane. Preferred sequences of this type are set forth in SEQ ID NO: 5 and 6.

SEQ ID NO: 5 is a partial cDNA which includes the end of the coding sequence from nucleotide position 1 to 43, the 3' UTR from nucleotide position 44 to 279 and a poly(A) tail (FIG. 16). SEQ ID NO: 5 is homologous to SEQ ID NO: 6.

SEQ ID NO: 6 is homologous to SEQ ID NO: 5 and contains a 5' UTR from nucleotide position 1 to 91, a coding sequence from nucleotide position 92 to 349, a 3' UTR from nucleotide position 350 to 551 and a poly(A) tail (FIG. 17). The deduced amino acid sequences from SEQ ID NO: 5 and 6 (FIG. 16, SEQ ID NO: 22; FIG. 17, SEQ ID NO: 23) correspond to a ribosomal protein S27, with numerous examples of homologues found in the databases (examples can be found under accession numbers D231399 for a rice mRNA sequence, X85544 for a barley mRNA or L19739 for a human mRNA).

SEQ ID NO: 22 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 5 (FIG. 16) (see SEQ ID NO: 5).

SEQ ID NO: 23 is the amino acid sequence deduced from the ORF in the nucleotide sequence corresponding to SEQ ID NO: 6 (FIG. 17) (see SEQ ID NO: 6).

2.4. Polynucleotide sequence variants of transcribed DNA sequences

In one embodiment, isolated nucleic acid variants of the invention may be prepared according to the following procedure:

(i) obtaining a nucleic acid extract from a suitable tissue of a plant, preferably a monocotyledonous plant;

(ii) creating primers which are optionally degenerate wherein each comprises a portion of a transcribable DNA sequence of the invention; and

(iii) using said primers to amplify, via nucleic acid amplification techniques, at least one amplification product from said nucleic acid extract, wherein said amplification product corresponds to a polynucleotide sequence variant.

Suitable nucleic acid amplification techniques are well known to the skilled addressee, and include polymerase chain reaction (PCR) as for example described in Ausubel *et al.* (*supra*) which is incorporated herein by reference; strand displacement amplification (SDA) as for example described in U.S. Patent No 5,422,252 which is incorporated herein by reference; rolling circle replication (RCR) as for example described in Liu *et al.*, (1996, *J. Am. Chem. Soc.* 118:1587-1594 and International application WO 92/01813) and Lizardi *et al.*, (International Application WO 97/19193) which are incorporated herein by reference; nucleic acid sequence-based amplification (NASBA) as for example described by Sooknanan *et al.*, (1994, *Biotechniques* 17:1077-1080) which is incorporated herein by reference; and Q-β replicase amplification as for example described by Tyagi *et al.*, (1996, *Proc. Natl. Acad. Sci. USA* 93:5395-5400) which is incorporated herein by reference.

An embodiment if this method emplying inverse PCR and primers according to SEQ ID NOS 24 and 25 will be described hereinafter.

Typically, polynucleotide sequence variants that are substantially complementary to a reference polynucleotide are identified by blotting techniques that include a step whereby nucleic acids are immobilised on a matrix (preferably a synthetic membrane such as nitrocellulose), followed by a hybridisation step, and a detection step. Southern blotting is used to identify a complementary DNA sequence; northern blotting is used to identify a complementary RNA sequence. Dot blotting and slot blotting can be used to identify complementary DNA/DNA, DNA/RNA or RNA/RNA polynucleotide sequences. Such techniques are well known by those skilled in the art, and have been described in Ausubel *et al.* (1994-1998, *supra*) at pages 2.9.1 through 2.9.20.

According to such methods, Southern blotting involves separating DNA molecules according to size by gel electrophoresis, transferring the size-separated DNA to a synthetic membrane, and hybridising the membrane-bound DNA to a complementary nucleotide sequence labelled radioactively, enzymatically or fluorochromatically. In dot blotting and slot blotting, DNA samples are directly applied to a synthetic membrane prior to hybridisation as above.

An alternative blotting step is used when identifying complementary

polynucleotides in a cDNA or genomic DNA library, such as through the process of plaque or colony hybridisation. A typical example of this procedure is described in Sambrook *et al.* ("Molecular Cloning. A Laboratory Manual", Cold Spring Harbour Press, 1989) Chapters 8-12.

Typically, the following general procedure can be used to determine hybridisation conditions. Polynucleotides are blotted/transferred to a synthetic membrane, as described above. A reference polynucleotide such as a polynucleotide of the invention is labelled as described above, and the ability of this labelled polynucleotide to hybridise with an immobilised polynucleotide is analysed.

A skilled addressee will recognize that a number of factors influence hybridisation. The specific activity of radioactively labelled polynucleotide sequence should typically be greater than or equal to about 10^8 dpm/mg to provide a detectable signal. A radiolabelled nucleotide sequence of specific activity 10^8 to 10^9 dpm/mg can detect approximately 0.5 pg of DNA. It is well known in the art that sufficient DNA must be immobilised on the membrane to permit detection. It is desirable to have excess immobilised DNA, usually 10 µg. Adding an inert polymer such as 10% (w/v) dextran sulfate (MW 500,000) or polyethylene glycol 6000 during hybridisation can also increase the sensitivity of hybridisation (see Ausubel *supra* at 2.10.10).

To achieve meaningful results from hybridisation between a polynucleotide immobilised on a membrane and a labelled polynucleotide, a sufficient amount of the labelled polynucleotide must be hybridised to the immobilized polynucleotide following washing. Washing ensures that the labelled polynucleotide is hybridised only to the immobilized polynucleotide with a desired degree of complementarity to the labelled polynucleotide.

It will be understood that polynucleotide sequence variants according to the invention will hybridise to a reference polynucleotide under at least low stringency conditions. Reference herein to low stringency conditions include and encompass from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridisation at 42°C, and at least about 1 M to at least about 2 M salt for washing at 42°C. Low stringency conditions

also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65°C, and (i) 2xSSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at room temperature.

Preferably, the polynucleotide sequence variants hybridise to a reference polynucleotide under at least medium stringency conditions. Medium stringency conditions include and encompass from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridisation at 42°C, and at least about 0.5 M to at least about 0.9 M salt for washing at 42°C. Medium stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65°C, and (i) 2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at 42°C.

More preferably, the polynucleotide sequence variants hybridise to a reference polynucleotide under high stringency conditions. High stringency conditions include and encompass from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01 M to at least about 0.15 M salt for hybridisation at 42°C, and at least about 0.01 M to at least about 0.15 M salt for washing at 42°C. High stringency conditions also may include 1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65°C, and (i) 0.2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 1% SDS for washing at a temperature in excess of 65°C.

Other stringent conditions are well known in the art. A skilled addressee will recognize that various factors can be manipulated to optimize the specificity of the hybridisation. Optimisation of the stringency of the final washes can serve to ensure a high degree of hybridisation. For detailed examples, see Ausubel *et al.*, *supra* at pages 2.10.1 to 2.10.16 and Sambrook *et al.* (1989, *supra*) at sections 1.101 to 1.104, which are incorporated herein by reference.

While stringent washes are typically carried out at temperatures from about 42°C to 68°C, one skilled in the art will appreciate that other temperatures may be suitable for stringent conditions. Maximum hybridisation typically occurs at about 20°C to 25°C below the T_m for formation of a DNA-DNA hybrid. It is well known in

the art that the T_m is the melting temperature, or temperature at which two complementary polynucleotide sequences dissociate. Methods for estimating T_m are well known in the art (see Ausubel *et al.*, *supra* at page 2.10.8).

In general, the T_m of a duplex DNA decreases by about 1°C with every increase of 1% in the number of mismatched base pairs.

In a preferred hybridisation procedure, a membrane (*e.g.*, a nitrocellulose membrane or a nylon membrane) containing immobilised DNA is hybridised overnight at 42°C in a hybridisation buffer (50% deionised formamide, 5xSSC, 5x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrollidone and 0.1% bovine serum albumin), 0.1% SDS and 200 mg/mL denatured salmon sperm DNA) containing labelled probe. The membrane is then subjected to two sequential medium stringency washes (*i.e.*, 2xSSC/0.1% SDS for 15 min at 45°C, followed by 2xSSC/0.1% SDS for 15 min at 50°C), followed by two sequential high stringency washes (*i.e.*, 0.2xSSC/0.1% SDS for 12min at 55°C followed by 0.2xSSC and 0.1%SDS solution for 12 min).

Methods for detecting a labelled polynucleotide hybridised to an immobilised polynucleotide are well known to practitioners in the art. Such methods include autoradiography, phosphorimaging, chemiluminescent, fluorescent and colorimetric detection.

3. Promoter sequences of the invention

3.1. Promoters regions of specific transcribed DNA sequences

The invention also provides promoter regions isolated adjacent to the start of the transcribed DNA sequences described in Section 2. In particular, stem-specific and constitutive promoters for expression of chimeric or heterologous genes in plants, preferably monocotyledonous plants are provided. Preferred stem-specific promoters of the invention may be selected from the polynucleotide sequences set forth in SEQ ID NO: 12 or SEQ ID NO: 13. Preferred constitutive promoters of the invention include the polynucleotides set forth in SEQ ID NO: 14 or SEQ ID NO: 15.

The invention also contemplates biologically-active fragments of any

one of SEQ ID NO: 12, 13, 14 or 15 as well as polynucleotide sequence variants thereof. Those of skill in the art will understand that a biologically-active fragment of a promoter sequence, when fused to a particular gene and introduced into a plant cell, causes expression of the gene at a level higher than is possible in the absence of such fragment. The activity of a promoter can be determined by methods well known in the art. For example, reference may be made to Medberry *et al.* (1992, *Plant Cell* 4:185; 1993, *The Plant J.* 3:619, incorporated herein by reference), Sambrook *et al.* (1989, *supra*) and McPherson *et al.* (U.S. Patent No. 5,164,316, incorporated herein by reference).

3.2. Promoter variants

Promoter variants that are substantially complementary to a reference promoter of the invention may be obtained by procedures outlined in Section 2.3.

In general, variants comprise regions that show at least 70%, more suitably at least 80%, preferably at least 90%, and most preferably at least 95% sequence identity over a reference promoter sequence of identical size ("*comparison window*") or when compared to an aligned sequence in which the alignment is performed by a computer homology program known in the art. What constitutes suitable variants may be determined by conventional techniques. For example, polynucleotides according to SEQ ID NO: 12, 13, 14 and 15 can be mutated using random mutagenesis (*e.g.*, transposon mutagenesis), oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis and cassette mutagenesis of an earlier prepared variant or non-variant version of an isolated natural promoter according to the invention.

Oligonucleotide-mediated mutagenesis is a preferred method for preparing nucleotide substitution variants of a promoter of the invention. This technique is well known in the art as, for example, described by Adelman *et al.* (1983, *DNA* 2:183). Briefly, promoter DNA is altered by hybridising an oligonucleotide encoding the desired mutation to a template DNA, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the promoter of interest. After hybridisation, a DNA polymerase is used to synthesise an entire second complementary strand of the

template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the promoter of interest.

Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridise properly to the single-stranded DNA template molecule.

The DNA template can be generated by those vectors that are either derived from bacteriophage M13 vectors, or those vectors that contain a single-stranded phage origin of replication as described by Viera *et al.* (1987, *Methods Enzymol.* 153:3). Thus, the DNA that is to be mutated may be inserted into one of the vectors to generate single-stranded template. Production of single-stranded template is described, for example, in Sections 4.21-4.41 of Sambrook *et al.* (1989, *supra*).

Alternatively, the single-stranded template may be generated by denaturing double-stranded plasmid (or other DNA) using standard techniques.

For alteration of the native DNA sequence, the oligonucleotide is hybridised to the single-stranded template under suitable hybridisation conditions. A DNA polymerising enzyme, usually the Klenow fragment of DNA polymerase I, is then added to synthesise the complementary strand of the template using the oligonucleotide as a primer for synthesis. A heteroduplex molecule is thus formed such that one strand of DNA encodes the mutated form of the promoter under test, and the other strand (the original template) encodes the native unaltered sequence of the promoter under test. This heteroduplex molecule is then transformed into a suitable host cell, usually a prokaryote such as *E. coli*. After the cells are grown, they are plated onto agarose plates and screened using the oligonucleotide primer having a detectable label to identify the bacterial colonies having the mutated DNA. The resultant mutated DNA fragments are then cloned into suitable expression hosts such as *E. coli* using conventional technology and clones that retain the desired promoter activity are detected. Where the clones have been derived using random mutagenesis techniques, positive clones would have to be sequenced in order to detect the

mutation.

Alternatively, linker-scanning mutagenesis of DNA may be used to introduce clusters of point mutations throughout a sequence of interest that has been cloned into a plasmid vector. For example, reference may be made to Ausubel *et al.*, *supra*, (in particular, Chapter 8.4, incorporated herein by reference) which describes a first protocol that uses complementary oligonucleotides and requires a unique restriction site adjacent to the region that is to be mutagenised. A nested series of deletion mutations is first generated in the region. A pair of complementary oligonucleotides is synthesised to fill in the gap in the sequence of interest between the linker at the deletion endpoint and the nearby restriction site. The linker sequence actually provides the desired clusters of point mutations as it is moved or "scanned" across the region by its position at the varied endpoints of the deletion mutation series. An alternate protocol is also described by Ausubel *et al.*, *supra*, which makes use of site directed mutagenesis procedures to introduce small clusters of point mutations throughout the target region. Briefly, mutations are introduced into a sequence by annealing a synthetic oligonucleotide containing one or more mismatches to the sequence of interest cloned into a single-stranded M13 vector. This template is grown in an *Escherichia coli dut⁻ ung⁻* strain, which allows the incorporation of uracil into the template strand. The oligonucleotide is annealed to the template and extended with T4 DNA polymerase to create a double-stranded heteroduplex. Finally, the heteroduplex is introduced into a wild-type *E. coli* strain, which will prevent replication of the template strand due to the presence of apurinic sites (generated where uracil is incorporated), thereby resulting in plaques containing only mutated DNA.

Region-specific mutagenesis and directed mutagenesis using PCR may also be employed to construct promoter variants according to the invention. In this regard, reference may be made, for example, to Ausubel *et al.*, *supra*, in particular Chapters 8.2A and 8.5, which are incorporated herein by reference.

4. <u>Chimeric DNA constructs and expression vectors</u>

An isolated nucleic acid promoter or variant according to the invention can be fused to a heterologous nucleic acid to form a chimeric gene. The

heterologous nucleic acid may be a foreign or endogenous DNA sequence. For the purposes of transformation and expression of the heterologous nucleic acid in plants, it is preferred that the chimeric gene includes regulatory sequences which influence expression of the heterologous nucleic acid in plants. Preferably, the chimeric gene is present in an expression vector which includes regulatory sequences that enable selective propagation in bacteria.

4.1 3' Non-translated region

A 3' non-translated sequence refers to that portion of a gene comprising a DNA segment that contains a polyadenylation signal and any other regulatory signals capable of effecting mRNA processing or gene expression. The polyadenylation signal is characterized by effecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. Polyadenylation signals are commonly recognized by the presence of homology to the canonical form 5' AATAAA-3' although variations are not uncommon.

The 3' non-translated regulatory DNA sequence preferably includes from about 50 to 1,000 nucleotide base pairs and contains plant transcriptional and translational termination sequences. Examples of suitable 3' non-translated sequences are the 3' transcribed non-translated regions containing a polyadenylation signal from the nopaline synthase (*nos*) gene of *Agrobacterium tumefaciens* (Bevan *et al.*, 1983, *Nucl. Acid Res.*, 11:369) and the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*. Alternatively, suitable 3' non-translated sequences may be derived from plant genes such as the 3' end of the protease inhibitor I or II genes from potato or tomato, the soybean storage protein genes and the pea E9 small subunit of the ribulose-1,5-bisphosphate carboxylase (ssRUBISCO) gene, although other 3' elements known to those of skill in the art can also be employed. Alternatively, 3' non-translated regulatory sequences can be obtained *de novo* as, for example, described by An (1987, *Methods in Enzymology*, 153:292), which is incorporated herein by reference.

4.2 Optional sequences

The chimeric DNA construct of the present invention can further include enhancers, either translation or transcription enhancers, as may be required. These enhancer regions are well known to persons skilled in the art, and can include the ATG initiation codon and adjacent sequences. The initiation codon must be in phase with the reading frame of the coding sequence relating to the foreign or endogenous DNA sequence to ensure translation of the entire sequence. The translation control signals and initiation codons can be of a variety of origins, both natural and synthetic. Translational initiation regions may be provided from the source of the transcriptional initiation region, or from the foreign or endogenous DNA sequence. The sequence can also be derived from the source of the promoter selected to drive transcription, and can be specifically modified so as to increase translation of the mRNA.

Examples of transcriptional enhancers include, but are not restricted to, elements from the CaMV 35S promoter and octopine synthase genes as for example described by Last *et al.* (U.S. Patent No. 5,290,924, which is incorporated herein by reference). It is proposed that the use of an enhancer element such as the *ocs* element, and particularly multiple copies of the element, will act to increase the level of transcription from adjacent promoters when applied in the context of plant transformation.

As the DNA sequence inserted between the transcription initiation site and the start of the coding sequence, *i.e.*, the untranslated leader sequence, can influence gene expression, one can also employ a particular leader sequence. Preferred leader sequences include those that comprise sequences selected to direct optimum expression of the foreign or endogenous DNA sequence. For example, such leader sequences include a preferred consensus sequence which can increase or maintain mRNA stability and prevent inappropriate initiation of translation as for example described by Joshi (1987, *Nucl. Acid Res.*, 15:6643), which is incorporated herein by reference. However, other leader sequences, *e.g.*, the leader sequence of RTBV, have a high degree of secondary structure that is expected to decrease mRNA stability and/or decrease translation of the mRNA. Thus, leader sequences (i) that do not have a high degree of secondary structure, (ii) that have a high degree of

secondary structure where the secondary structure does not inhibit mRNA stability and/or decrease translation, or (iii) that are derived from genes that are highly expressed in plants, will be most preferred.

Regulatory elements such as the sucrose synthase intron as, for example, described by Vasil *et al.* (1989, *Plant Physiol.*, 91:5175), the Adh intron I as, for example, described by Callis *et al.* (1987, *Genes Develop.*, II), or the TMV omega element as, for example, described by Gallie *et al.* (1989, *The Plant Cell*, 1:301) can also be included where desired. Other such regulatory elements useful in the practice of the invention are known to those of skill in the art.

Additionally, targeting sequences may be employed to target a protein product of the foreign or endogenous DNA sequence to an intracellular compartment within plant cells or to the extracellular environment. For example, a DNA sequence encoding a transit or signal peptide sequence may be operably linked to a sequence encoding a desired protein such that, when translated, the transit or signal peptide can transport the protein to a particular intracellular or extracellular destination, respectively, and can then be post-translationally removed. Transit or signal peptides act by facilitating the transport of proteins through intracellular membranes, *e.g.*, vacuole, vesicle, plastid and mitochondrial membranes, whereas signal peptides direct proteins through the extracellular membrane. For example, the transit or signal peptide can direct a desired protein to a particular organelle such as a plastid (*e.g.*, a chloroplast), rather than to the cytoplasm. Thus, the chimeric DNA construct can further comprise a plastid transit peptide encoding DNA sequence operably linked between a promoter region or promoter variant according to the invention and the foreign or endogenous DNA sequence. For example, reference may be made to Heijne *et al.* (1989, *Eur. J. Biochem.*, 180:535) and Keegstra *et al.* (1989, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 40:471), which are incorporated herein by reference.

A chimeric DNA construct can also be introduced into a vector, such as a plasmid. Plasmid vectors include additional DNA sequences that provide for easy selection, amplification, and transformation of the expression cassette in prokaryotic and eukaryotic cells, *e.g.*, pUC-derived vectors, pSK-derived vectors, pGEM-derived vectors, pSP-derived vectors, or pBS-derived vectors. Additional DNA sequences include origins of replication to provide for autonomous replication of the vector,

selectable marker genes, preferably encoding antibiotic or herbicide resistance, unique multiple cloning sites providing for multiple sites to insert DNA sequences or genes encoded in the chimeric DNA construct, and sequences that enhance transformation of prokaryotic and eukaryotic cells.

The vector preferably contains an element(s) that permits stable integration of the vector into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell. The vector may be integrated into the host cell genome when introduced into a host cell. For integration, the vector may rely on the foreign or endogenous DNA sequence or any other element of the vector for stable integration of the vector into the genome by homologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location in the chromosome. To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAM.beta.1 permitting replication in *Bacillus*. The origin of replication may be one having a mutation to make its function temperature-sensitive in a *Bacillus* cell (see, *e.g.*, Ehrlich, 1978, *Proc. Natl. Acad. Sci. USA* 75:1433).

4.3 Marker genes

To facilitate identification of transformants, the chimeric DNA

construct desirably comprises a selectable or screenable marker gene as, or in addition to, the expressible foreign or endogenous DNA sequence. The actual choice of a marker is not crucial as long as it is functional (*i.e.*, selective) in combination with the plant cells of choice. The marker gene and the foreign or endogenous DNA sequence of interest do not have to be linked, since co-transformation of unlinked genes as, for example, described in U.S. Pat. No. 4,399,216 is also an efficient process in plant transformation.

Included within the terms selectable or screenable marker genes are genes that encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers that encode a secretable antigen that can be identified by antibody interaction, or secretable enzymes that can be detected by their catalytic activity. Secretable proteins include, but are not restricted to, proteins that are inserted or trapped in the cell wall (*e.g.*, proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S); small, diffusible proteins detectable, *e.g.* by ELISA; and small active enzymes detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin acetyltransferase).

4.3.1 Selectable markers

Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers that confer antibiotic resistance such as ampicillin, kanamycin, erythromycin, chloramphenicol or tetracycline resistance. Exemplary selectable markers for selection of plant transformants include, but are not limited to, a *hyg* gene which encodes hygromycin B resistance; a neomycin phosphotransferase (*neo*) gene conferring resistance to kanamycin, paromomycin, G418 and the like as, for example, described by Potrykus *et al.* (1985, *Mol. Gen. Genet.* 199:183); a glutathione-S-transferase gene from rat liver conferring resistance to glutathione derived herbicides as, for example, described in EP-A 256 223; a glutamine synthetase gene conferring, upon overexpression, resistance to glutamine synthetase inhibitors such as phosphinothricin as, for example, described WO87/05327, an acetyl transferase gene from *Streptomyces viridochromogenes* conferring resistance to the selective agent phosphinothricin as,

for example, described in EP-A 275 957, a gene encoding a 5-enolshikimate-3-phosphate synthase (EPSPS) conferring tolerance to N-phosphonomethylglycine as, for example, described by Hinchee *et al.* (1988, *Biotech.*, 6:915), a *bar* gene conferring resistance against bialaphos as, for example, described in WO91/02071; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to bromoxynil (Stalker *et al.*, 1988, *Science*, 242:419); a dihydrofolate reductase (DHFR) gene conferring resistance to methotrexate (Thillet *et al.*, 1988, *J. Biol. Chem.*, 263:12500); a mutant acetolactate synthase gene (ALS), which confers resistance to imidazolinone, sulfonylurea or other ALS-inhibiting chemicals (EP-A-154 204); a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan; or a dalapon dehalogenase gene that confers resistance to the herbicide.

4.3.2 Screenable markers

Preferred screenable markers include, but are not limited to, a *uidA* gene encoding a β-glucuronidase (GUS) enzyme for which various chromogenic substrates are known; a β-galactosidase gene encoding an enzyme for which chromogenic substrates are known; an aequorin gene (Prasher *et al.*, 1985, *Biochem. Biophys. Res. Comm.*, 126:1259), which may be employed in calcium-sensitive bioluminescence detection; a green fluorescent protein gene (Niedz *et al.*, 1995 *Plant Cell Reports*, 14:403); a luciferase (*luc*) gene (Ow *et al.*, 1986, *Science*, 234:856), which allows for bioluminescence detection; a β-lactamase gene (Sutcliffe, 1978, *Proc. Natl. Acad. Sci. USA* 75:3737), which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); an R-locus gene, encoding a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, 1988, in *Chromosome Structure and Function*, pp. 263-282); an α-amylase gene (Ikuta *et al.*, 1990, *Biotech.*, 8:241); a tyrosinase gene (Katz *et al.*, 1983, *J. Gen. Microbiol.*, 129:2703) which encodes an enzyme capable of oxidizing tyrosine to dopa and dopaquinone which in turn condenses to form the easily detectable compound melanin; or a *xylE* gene (Zukowsky *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80:1101), which encodes a catechol dioxygenase that can convert chromogenic catechols.

5. Uses of the promoters of the invention

The isolated nucleic acid promoters of the invention may be used, *inter alia*, to drive expression of a foreign or endogenous DNA sequence. Preferred agronomic properties encoded by the foreign or endogenous DNA sequence include, but are not limited to, insect resistance or tolerance, herbicide resistance or tolerance, disease resistance or tolerance (*e.g.*, resistance to viruses or fungal pathogens), stress tolerance (increased salt tolerance) and improved food content or increased yields.

The foreign or endogenous DNA sequence may comprise a region transcribed into a molecule that modulates the expression of a corresponding target gene. The molecule may be an antisense RNA or a ribozyme or other transcript aimed at downregulation of expression of the corresponding target gene.

Anti-sense regulation and the use of ribozymes and co-suppression in plants are well known in the art. However, the skilled person is referred to United States Patent 5,759,829 for an example of antisense technology and to U.S. patent 5,707,835, U.S. patent 5,747,335 and U.S. patent 5,840,874 which each provide examples of ribozyme technology. With regard to co-suppression, reference is made to U.S. patent 5,283,184, U.S. patent 5,686,649 and International Publication WO98/53083 for examples of this technology. Each of these patent documents is incorporated herein by reference.

Alternatively, the foreign or endogenous DNA sequence may encode a molecule which is readily detectable or measurable, *e.g.* β-glucuronidase or luciferase; a selectable product, *e.g.*, neomycin phosphotransferase (nptII) conferring resistance to aminoglycosidic antibiotics such as geneticin and paramomycin; a product conferring herbicide tolerance, *e.g.* glyphosate resistance or glufosinate resistance; a product affecting starch biosynthesis or modification *e.g.* starch branching enzyme, starch synthases, ADP-glucose pyrophosphorylase; a product involved in fatty acid biosynthesis, *e.g.* desaturase or hydroxylase; a product conferring insect resistance, *e.g.* crystal toxin protein of *Bacillus thuringiensis*; a product conferring viral resistance, *e.g.* viral coat protein; a product conferring fungal resistance, *e.g.* chitinase, β-1,3-glucanase or phytoalexin; a product altering sucrose metabolism, *e.g.* invertase or sucrose synthase; a gene encoding valuable

pharmaceuticals, *e.g.* antibiotics, secondary metabolites, pharmaceutical peptides or vaccines.

6. Introduction of isolated nucleic acids into plant cells

A number of techniques are available for the introduction of DNA into a plant host cell. There are many plant transformation techniques well known to workers in the art, and new techniques are continually becoming known. The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practising the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce a chimeric DNA construct into plant cells is not essential to or a limitation of the invention, provided it achieves an acceptable level of nucleic acid transfer. Guidance in the practical implementation of transformation systems for plant improvement is provided by Birch (1997, *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **48**: 297-326), which is incorporated herein by reference.

In principle both dicotyledonous and monocotyledonous plants that are amenable to transformation, can be modified by introducing a chimeric DNA construct according to the invention into a recipient cell and growing a new plant that harbors and expresses the foreign or endogenous DNA sequence.

Introduction and expression of foreign or chimeric DNA sequences in dicotyledonous (broadleafed) plants such as tobacco, potato and alfalfa has been shown to be possible using the T-DNA of the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* (See, for example, Umbeck, U.S. Patent No. 5,004,863, and International application PCT/US93/02480). A construct of the invention may be introduced into a plant cell utilizing *A. tumefaciens* containing the Ti plasmid. In using an *A. tumefaciens* culture as a transformation vehicle, it is most advantageous to use a non-oncogenic strain of the *Agrobacterium* as the vector carrier so that normal non-oncogenic differentiation of the transformed tissues is possible. It is preferred that the *Agrobacterium* harbors a binary Ti plasmid system. Such a binary system comprises (1) a first Ti plasmid having a virulence region essential for the introduction of transfer DNA (T-DNA) into plants, and (2) a chimeric plasmid. The

chimeric plasmid contains at least one border region of the T-DNA region of a wild-type Ti plasmid flanking the nucleic acid to be transferred. Binary Ti plasmid systems have been shown effective to transform plant cells as, for example, described by De Framond (1983, *Biotechnology*, 1:262) and Hoekema *et al.* (1983, *Nature*, 303:179). Such a binary system is preferred *inter alia* because it does not require integration into the Ti plasmid in *Agrobacterium*.

Methods involving the use of *Agrobacterium* include, but are not limited to: (a) co-cultivation of *Agrobacterium* with cultured isolated protoplasts; (b) transformation of plant cells or tissues with *Agrobacterium*; or (c) transformation of seeds, apices or meristems with *Agrobacterium*.

Recently, rice, corn, pineapple and sugarcane, which are monocots, have been shown to be susceptible to transformation by *Agrobacterium*, for example as described in United States Patent No. 6,037,522, International Publication WO99/36637 and Arencibia *et al.* (1998, *Transgenic Res.* 7:213). However, some monocot crop plants have not yet been successfully transformed using *Agrobacterium*-mediated transformation. The Ti plasmid, however, may be manipulated in the future to act as a vector for these other monocot plants. Additionally, using the Ti plasmid as a model system, it may be possible to artificially construct transformation vectors for these plants. Ti plasmids might also be introduced into monocot plants by artificial methods such as microinjection, or fusion between monocot protoplasts and bacterial spheroplasts containing the T-region, which can then be integrated into the plant nuclear DNA.

In addition, gene transfer can be accomplished by *in situ* transformation by *Agrobacterium*, as described by Bechtold *et al.* (1993, *C.R. Acad. Sci. Paris*, 316:1194). This approach is based on the vacuum infiltration of a suspension of *Agrobacterium* cells.

Alternatively, nucleic acids may be introduced using root-inducing (Ri) plasmids of *Agrobacterium* as vectors.

Cauliflower mosaic virus (CaMV) may also be used as a vector for introducing of exogenous nucleic acids into plant cells (U.S. Pat. No. 4,407,956). CaMV DNA genome is inserted into a parent bacterial plasmid creating a

recombinant DNA molecule that can be propagated in bacteria. After cloning, the recombinant plasmid again may be cloned and further modified by introduction of the desired nucleic acid sequence. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the plant cells or plants.

Nucleic acids can also be introduced into plant cells by electroporation as, for example, described by Fromm *et al.* (1985, *Proc. Natl. Acad. Sci., U.S.A*, 82:5824) and Shimamoto *et al.* (1989, *Nature* 338:274-276). In this technique, plant protoplasts are electroporated in the presence of vectors or nucleic acids containing the relevant nucleic acid sequences. Electrical impulses of high field strength reversibly permeabilise membranes allowing the introduction of nucleic acids. Electroporated plant protoplasts reform the cell wall, divide and form a plant callus.

Another method for introducing nucleic acids into a plant cell is high velocity ballistic penetration by small particles (also known as particle bombardment or microprojectile bombardment) with the nucleic acid to be introduced contained either within the matrix of small beads or particles, or on the surface thereof as, for example described by Klein *et al.* (1987, *Nature* 327:70). Although typically only a single introduction of a new nucleic acid sequence is required, this method particularly provides for multiple introductions.

Alternatively, nucleic acids can be introduced into a plant cell by contacting the plant cell using mechanical or chemical means. For example, a nucleic acid can be mechanically transferred by microinjection directly into plant cells by use of micropipettes. Alternatively, a nucleic acid may be transferred into the plant cell by using polyethylene glycol which forms a precipitation complex with genetic material that is taken up by the cell.

Also contemplated are silicon carbide or tungsten whiskers, for example as described in United States Patent No. 5,302,523.

There are a variety of methods known currently for transformation of monocotyledonous plants. Presently, preferred methods for transformation of monocots are microprojectile bombardment of explants or suspension cells, and

direct DNA uptake or electroporation as, for example, described by Shimamoto *et al.* (1989, *supra*). Transgenic maize plants have been obtained by introducing the *Streptomyces hygroscopicus bar* gene into embryogenic cells of a maize suspension culture by microprojectile bombardment (Gordon-Kamm, 1990, *Plant Cell*, 2:603-618). The introduction of genetic material into aleurone protoplasts of other monocotyledonous crops such as wheat and barley has been reported (Lee, 1989, *Plant Mol. Biol.* 13:21-30). Wheat plants have been regenerated from embryogenic suspension culture by selecting only the aged compact and nodular embryogenic callus tissues for the establishment of the embryogenic suspension cultures (Vasil, 1990, *Bio/Technol.* 8:429-434). The combination with transformation systems for these crops enables the application of the present invention to monocots. These methods may also be applied for the transformation and regeneration of dicots. Transgenic sugarcane plants have been regenerated from embryogenic callus as, for example, described by Bower et al. (1996, *Molecular Breeding* 2:239-249).

Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, *e.g.*, bombardment with *Agrobacterium* coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with *Agrobacterium* (EP-A-486233).

The patent and scientific publications referred to above are all incorporated herein by reference.

7. Production and characterisation of transgenic plants

7.1 Regeneration

The methods used to regenerate transformed cells into differentiated plants are not critical to this invention, and any method suitable for a target plant can be employed. Normally, a plant cell is regenerated to obtain a whole plant following a transformation process.

Regeneration from protoplasts varies from species to species of plants, but generally a suspension of protoplasts is first made. In certain species, embryo formation can then be induced from the protoplast suspension, to the stage of

ripening and germination as natural embryos. The culture media will generally contain various amino acids and hormones, necessary for growth and regeneration. Examples of hormones utilized include auxins and cytokinins. It is sometimes advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these variables are controlled, regeneration is reproducible. Regeneration also occurs from plant callus, explants, organs or parts. Transformation can be performed in the context of organ or plant part regeneration as, for example, described in *Methods in Enzymology*, Vol. 118 and Klee *et al.* (1987, *Annual Review of Plant Physiology*, 38:467), which are incorporated herein by reference. Utilizing the leaf disk-transformation-regeneration method of Horsch *et al.* (1985, *Science*, 227:1229, incorporated herein by reference), disks are cultured on selective media, followed by shoot formation in about 2-4 weeks. Shoots that develop are excised from calli and transplanted to appropriate root-inducing selective medium. Rooted plantlets are transplanted to soil as soon as possible after roots appear. The plantlets can be repotted as required, until reaching maturity.

In vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenotes is made and new varieties are obtained and propagated vegetatively for commercial use.

In seed propagated crops, the mature transgenic plants can be self-crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced foreign gene(s). These seeds can be grown to produce plants that would produce the selected phenotype, *e.g.*, early flowering.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells that have been transformed as described. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

It will be appreciated that the literature describes numerous techniques for regenerating specific plant types and more are continually becoming known. Those of ordinary skill in the art can refer to the literature for details and select suitable techniques without undue experimentation.

7.2. Characterization

To confirm the presence of the heterologous nucleic acid in the regenerating plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting and PCR; a protein expressed by the heterologous DNA may be analysed by western blotting, high performance liquid chromatography or ELISA (*e.g.*, nptII) as is well known in the art.

Examples of various methods applicable to characterization of transgenic plants are provided in Chapters 9 and 11 of PLANT MOLECULAR BIOLOGY A Laboratory Manual Ed. M.S. Clark (Springer-Verlag, Heidelberg, 1997), which chapters are herein incorporated by reference.

So that the invention may be understood and more detail, the skilled person is directed to the following non-limiting examples.

EXAMPLE 1

Construction of a tissue-specific cDNA library

Extraction of tissue-specific RNA

Total RNA was extracted essentially according to the method of Chirgwin *et al.* (1979, Biochemistry 18 (24):5294-5299), from 9-month old Pindar sugarcane plants field-grown at the BSES Pathology Farm at Eight Mile Plains in Brisbane, Queensland, Australia. Roots, stems and leaves were harvested from three separate plants randomly chosen in the field. The tissues were washed and stripped of dead material and immediately quick-frozen in liquid nitrogen. Five to ten grams of frozen tissue were ground to a fine powder in liquid nitrogen and 40 mL of a 4 M guanidine isocyanate solution were added. After thawing the mixture, 140 µL of β-

mercaptoethanol and 2 mL of a 10% sarkosyl solution were added and thoroughly mixed. The cell debris was pelleted by centrifugation. The supernatant was then filtered through one layer of Miracloth (Calbiochem) and gently poured over a 5.7 M caesium chloride cushion. The RNA was spun down for 18 hours in a SW27 rotor at 20°C (25,000 RPM). The supernatant was carefully removed from the tubes and the RNA pellet was resuspended in 400 µL of DEPC-treated distilled water. After a phenol/chloroform extraction (phenol/chloroform/isoamyl alcohol 24/24/2), the RNA was precipitated by addition of 1/10 volume of sodium citrate (3M, pH 5.2) and 2.5 volumes of ethanol. The quantity of extracted RNA was determined by spectrophotometry (OD260nm) and the quality visualised by agarose gel electrophoresis (Sambrook *et al.* 1989).

Isolation of tissue-specific mRNA

The PolyATtract® mRNA Isolation Systems (Promega) was used to separate the poly(A) RNAs from the non-poly(A) RNAs. The system uses a biotinylated oligo(dT) primer to hybridise to the poly(A) tail of the mature mRNAs. The duplexes are captured by streptavidin coupled to a paramagnetic particle and isolated using a magnetic stand. The mRNAs are then eluted directly by DEPC-treated dH_2O. The yields of mRNA from total RNA were 0.37% (leaf), 0.31% (root) and 0.27% (stem).

cDNA synthesis

Stem mRNAs (5µg) were used in the synthesis of double-stranded cDNAs with the TimeSaver® cDNA Synthesis Kit (Pharmacia Biotech). The first-strand cDNA was synthesized using an oligo$(dT)_{12-18}$ primer. After synthesis of the second strand, *Eco*RI/*Not*I adaptors were ligated to the cDNAs.

cDNA library

A stem-specific cDNA library was constructed in Lambda ZAP®II (Predigested Lambda ZAP®II/*Eco*RI/CIAP Cloning Kit, Stratagene). Stem cDNAs (150 ng) were ligated to 1 µg of predigested Lambda ZAP®II vector for 16 hrs at 14°C. Half of the ligation was *in vitro*-packaged (Gigapack® II Gold Packaging

Extract, Stratagene). Titration of the library was carried out using 1/500 and 1/5000 of the phage-containing supernatant and XL1-Blue MRF' host bacteria. After the phage and the host cells were incubated together, top agar containing IPTG (2.5 mM) and X-gal (4 g/L) was added and the mixture plated on NZY solid medium plates. The titer was estimated at 715,000 white plaque forming units/µg of vector with 4.6% non-recombinant background plaques (blue plaques). The second half of the ligations was similarly packaged. The packaged phages were combined, the titer was estimated at 663,000 pfu/mL after allowing a 20% drop of the titer after a 4-day storage at 4°C. The cDNA library was amplified by plating 50,000 pfu/ 150 mm-NZY plate (six plates). After 8hrs incubation at 37°C, when the plaques were no more than 1 to 2 mm in diameter, the plates were overlaid with 8-10 mL of SM buffer (10 mM NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris-HCl pH 7.5 and 0.01% gelatin) and gently rocked overnight at 4°C. The bacteriophage suspension was recovered from each plate and pooled. Chloroform was added (5% final concentration) and the cell debris pelleted by centrifugation (10 min at 2000x*g*). The supernatant was recovered and chloroform added to 0.3% final concentration. Aliquots were stored in 7% DMSO at -80°C. The titer of the amplified library was 3 x 10^{10} pfu/mL.

EXAMPLE 2

Differential screening of the stem cDNA library, isolation of Lambda clones

Plating of the cDNA library

The stem cDNA library was plated with *E.coli* host strain LE392, at a density of 8,000 pfu/150 mm NZY plates (supplemented with 0.2% maltose). Ten plates were incubated at 37°C for 9 hrs, when the plaques reached a size of 1 mm or less in diameter without being confluent. Plaque growth was stopped by transferring the plates to 4°C. The phage plaques were transferred onto nylon membranes (Nylon N+, Amersham) in triplicates: the first lift was in contact with the plaques for 45 seconds, the second lift for 2 min and the third lift for 4 min. The phage DNA was then denatured by placing the membrane for 2 min on a 3MM chromatography paper (Whatman) saturated with denaturing solution (2M NaOH, 1.5M NaCl). The

membrane was then transferred onto a 3MM chromatography paper saturated with a neutralising buffer (0.5 M Tris-HCl pH 8.0, 1.5 M NaCl) for 5min and finally rinsed in a 0.2 M Tris-HCl pH 8.0 and 2xSSC buffer (0.2 M NaCl, 0.03 M sodium citrate, pH 7.0) for 1min. The phage DNA was then UV-cross-linked to the damp membranes (GS Gene Linker™ UV Chamber, Bio-Rad). The membranes were air-dried and stored until further use between sheets of 3MM chromatography papers.

Single-stranded radiolabelled cDNAs (probes)

Labelling reactions (30 µL total volume) contained 2 µg of mRNAs used as templates, 1 µL of Rnasin® (Promega), 6 µL of 5x buffer (250 mM Tris-HCl pH 8.3, 375 mM KCL, 15 mM $MgCl_2$), 2 µL of dNTPs solution (10 mM each of dATP, dTTP and dGTP), 2 µL of oligo(dT) (0.5 µg/µL), 5 µL of ^{32}PdCTP (10 µCi/µL, Amersham) and 1 µL of M-MLV Reverse Transcriptase (Life Technologies). The mRNAs were brought to 13 µL with DEPC-treated water and denatured for 10 min at 65°C prior to the addition of the other components. The reaction was carried out at 37°C for 1 hr. The probes were purified through a polyacrylamide column (Biospin 30 column, Bio-Rad).

Differential hybridisation

The nylon membranes were pre-hybridised for 6 hrs at 42°C in a pre-hybridisation solution containing 50% deionised formamide, 5xSSC (0.5 M NaCl, 0.075 M sodium citrate, pH 7.0), 5x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrollidone and 0.1% bovine serum albumin), 0.1% sodium dodecyl sulfate (SDS) and 200 mg/mL denatured salmon sperm DNA. The radiolabelled probe was added directly to the pre-hybridisation solution after completion of the pre-hybridisation procedure. Sets of membranes were hybridised as follows: the first lifts were hybridised to the leaf radiolabelled cDNAs, the second lifts were hybridised to the root radiolabelled cDNAs and the third lifts were hybridised to the stem radiolabelled cDNAs.

Washing of the membranes and exposure to a X-ray film

After a 20 hr hybridisation, the membranes were washed sequentially in 2xSSC and 0.1% SDS solution for 15 min at 45°C, 2xSSC and 0.1% SDS solution for 15 min at 50°C, 0.2xSSC and 0.1% SDS solution for 12min at 55°C and 0.2xSSC and 0.1%SDS solution for 12 min at 65°C. The membranes were wrapped in a plastic film and exposed for two days to X-ray film (X-Omat Diagnostic Film, Kodak).

Primary screening

Identification of differentially expressed cDNA clones (stem-specific)

Each set of films corresponding to the same initial plate was visually examined for differentially hybridising plaques: for example, the film 1.1 (corresponding to the plate 1, first lift, probed with the leaf radiolabelled cDNAs) was overlaid with the film 1.3 (corresponding to the plate 1, third lift, probed with the stem radiolabelled cDNAs). The ten sets of films were screened for plaques hybridising to the stem radiolabelled cDNAs but giving no signals (or faint signals) when probed by either the leaf or root radiolabelled cDNAs. This method allowed the identification of 44 plaques corresponding to potentially stem-specific cDNA clones.

Identification of constitutively expressed cDNA clones

The same method revealed many plaques that strongly hybridised to each of the three different independently applied radiolabelled cDNAs (*i.e.*, stem, root and leaf). Seven plaques were selected amongst many that fit the requirements.

Secondary screening

Identification of differentially expressed cDNA clones (stem-specific)

The plaques identified during the primary screening were cored from the plates and eluted in 1 mL of SM buffer and 20 µL of chloroform. The titers were determined for each eluate. A top agar bacterial lawn was poured onto 100 mm-NZY plates (supplemented with 0.2% maltose). About 50 phages from each eluate were

spread onto 1/4 of the plate. Plaques were grown at 37°C for 8 hrs, then placed at 4°C. Triplicate lifts were made from each plate (Nylon N+, Amersham) as described above. Pre-hybridisation, hybridisation and visual screening, carried out in the same manner as for the primary screening, confirmed 37 plaques with a possible stem-specific expression pattern.

Identification of constitutively expressed cDNA clones

Secondary screening confirmed all of seven tested constitutive clones.

In vivo excision of the pBluescript SK(-) phagemid from the Lambda ZAP II vector

The excision of the phagemid from the Lambda ZAP II vector was performed as described in the manual of the Lambda ZAP® II/*Eco*RI/CIAP Cloning Kit (Stratagene). The ExAssist/SOLR system procedure results in SOLR colonies containing the pBluescript SK(-) double-stranded phagemid with a cloned DNA insert. The plasmids were isolated and restricted by either *Eco*RI or *Not*I to release the cloned cDNA which was purified from an agarose gel and used to generate radiolabelled probes by random priming for Southern and northern analyses.

EXAMPLE 3

Northern and Southern analyses of the cDNA clones

Northern analysis

The specificity of the expression of each clone was verified by northern analysis. Total RNA was extracted from different tissues (i.e. roots, different stem internodes and nodes, young emerging leaves, mature leaves and older leaves) of field-grown sugarcane and maize plants and glasshouse-grown sorghum plants as described in Example 1. The RNA sample was mixed with 6 volumes of RNA sample buffer (60% deionised formamide, 25% of 37% formaldehyde and 15% of a 10xMOPS buffer containing 0.2 M MOPS pH 7.0, 50 mM sodium citrate and 5 mM EDTA pH 8.0) and 1 volume of RNA loading solution (50% glycerol, 1mM EDTA pH 8.0, 2 mg/mL bromophenol blue and 2 mg/mL xylene cyanol). Prior to loading the gel, the samples were heat-denatured at 65°C for 10 min. Equal amounts

of total RNA (8 µg/lane) were loaded onto an agarose gel. The gel was run in 1xTBE buffer (0.1 M Tris, 0.09 M Boric acid and 1 mM EDTA). Transfer of the RNA to a nylon membrane (Nylon N+, Amersham) was by capillary blotting (Sambrook *et al.* 1989). Pre-hybridisation and hybridisation were carried out as described before, using probes labelled with ^{32}PdCTP (Amersham Rediprime DNA labelling system). Blots were washed at the stringency previously described, and scanned with a PhosphorImager SI (Molecular Dynamics). FIGS 1 and 2 show the expression pattern of the cDNA clone c67 (SEQ ID NO: 1) that is preferentially expressed in the stem of field-grown sugarcane plants. The relative expression levels are 100 in the stem, 0.3-0.7 in the leaves and between 1.7 and 5 in the roots (depending on the cultivar). Within the stem, c67 is predominantly expressed in the more mature parts of the stem. FIGS. 3, 4, 5 and 6 show the stem-specific expression pattern of the cDNA clone c51 (SEQ ID NO: 2, 3 and 4). The c51 probe also reveals a homologous stem-specific cDNA in maize and sorghum plants (with some expression in the roots of glasshouse-grown sorghum plants). Figures 7, 8, 9 and 10 show the constitutive expression of the cDNA clone c32A (SEQ ID NO: 5 and 6) within the different parts of the sugarcane, maize, sorghum and rice plants that have been tested.

Southern analysis

DNA was extracted from sugarcane cultivars Pindar or Q117, digested by selected restriction enzymes, and electrophoresed (10 µg/lane) on an agarose gel in 1xTAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.3). The DNA was then denatured for 1hr (2 M NaOH, 1.5 M NaCl), neutralised for 1hr (0.5 M Tris-HCl pH8.0, 1.5 M NaCl) and equilibrated in 10xSSC prior to the transfer of the DNA onto the nylon N+ membranes (Amersham) by capillary blotting. The membranes were then treated as for the northern blots for pre-hybridisation, hybridisation to a radiolabelled probe, washes and detection by PhosphorImager (Molecular Dynamics). Southern blot analyses of the clones c67 (SEQ ID NO: 1), c51 (SEQ ID NO:2, 3 and 4) and c322 (SEQ ID NO:5 and 6) are shown respectively in FIGS. 11, 12 and 13.

EXAMPLE 4

Screening of the stem-specific cDNA library

The stem-specific cDNA library was screened as described above, using c51 (SEQ ID NO: 2) and c32A (SEQ ID NO: 5) as probes. After the secondary screening, one additional clone corresponding to c32A and two additional clones, corresponding to c51, were isolated (c322, c511 and c512, corresponding respectively to SEQ ID NOS: 6, 3 and 4).

EXAMPLE 5

Sequencing of the clones

Plasmid inserts were sequenced using universal primers (T3, T7, SP6, M13 Reverse and M13 Forward) or custom-designed primers, and the ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit was used and the reactions loaded onto an ABI PRISM instrument. Sequences were obtained from both strands of the DNA.

The nucleotide sequence of the cDNA clone c67 (976nt) is shown in figure 14 (SEQ ID NO: 1). FIG. 15 shows the nucleotide sequence alignment of the homologous cDNA clones c51, c511 and c512 (SEQ ID NO: 2, 3 and 4). FIGS. 16 and 17 show the nucleotide sequences of the cDNA clones c32A and c322 respectively (SEQ ID NO: 5 and 6). FIGS. 23, 24, 26, 29 and 31 show the nucleotide sequences of cDNA clones c19 (SEQ ID NO: 7), c3 (SEQ ID NO: 8), c18 (SEQ ID NO: 9), c53A (SEQ ID NO: 10) and c57 (SEQ ID NO: 11), respectively. In addition, FIGS. 18, 19,20 and 21 show respectively the nucleotide sequences of the promoter sequences 67p (SEQ ID NO: 12), 32A2P2 (SEQ ID NO: 13), 32A6P2 (SEQ ID NO: 14) and 51p (SEQ ID NO: 15).

EXAMPLE 6

Promoter recovery by iPCR

Primer design for inverse PCR reaction

Two primers designed from the sequence of the cDNA clone c67 (SEQ ID NO: 1) for use in an inverse PCR reaction are 67-1 (5' AGGACGCTTCTCAGATTGGC 3'; SEQ ID NO:24) and 67-2 (5' GTATTCGGAGTTGCAGGTCG 3'; SEQ ID NO:25).

iPCR

Genomic DNA from sugarcane (cultivar Q117) was isolated through a CsCl gradient and restricted by *Sac*I. The enzyme was heat-inactivated (65°C for 10 min) and 600 ng of restricted DNA were recircularised in a total reaction volume of 400 µL containing 9 Weiss units of T4 DNA ligase (New England Biolabs). After completion of the ligation, the reaction was extracted with phenol/chloroform and the DNA precipitated by addition of ethanol. The PCR reaction was carried out in a final volume of 50 µL comprising 150 ng of circularised DNA, 10 µL of 5x buffer (for a final concentration of 60 mM Tris-SO_4 pH9.1, 18 mM $(NH_4)_2SO_4$ and 1.8mM $MgCl_2$), 20 ng of primer 67-1, 20 ng of primer 67-2, 5 µL of a dNTPs solution (2 mM each) and 1.5 µL of polymerase (ELONGASE™ Enzyme Mix, Life Technologies). After an initial 30 s at 94°C, 35 cycles were performed consisting of 1 min at 94°C, 1 min at 58°C and 10 min at 68°C. A second iPCR reaction was performed in the same conditions using 0.1 µL of the first reaction as template. Products of the iPCR reactions were run on an agarose gel and the DNA purified from the gel (Bresaclean, Bresatec). The purified DNA was blunt-ended using Klenow and restricted by *Sac*I. The two DNA products were cloned into pGEM®-4Z (Promega) digested by *Sma*I and *Sac*I. Sequencing using the T7 primer identified the plasmid clone containing a DNA region upstream of the primer 67-1 (plasmid pG4-67pro).

Fusion of the promoter region to a reporter gene

The full nucleotide sequence (to the ATG start codon) of the promoter region obtained by the inverse PCR method is shown in FIG. 18 (SEQ ID NO: 12). The GUS gene and *nos* terminator were excised from plasmid pBI101.3 (Clontech) as an *Xba*I-*Eco*RI fragment and ligated into the *Xba*I and *Eco*RI sites of pBluescript (Stratagene) to form plasmid pBS-GUS-3. A partial *Sac*I digest, followed by a *Bam*HI digest were carried out on the pBS-GUS-3 plasmid. Within these two sites was inserted a *Sac*I-*Bam*HI DNA fragment from the plasmid pG4-67pro to obtain plasmid p67G, with the GUS coding region translationally fused to the ATG start codon at the 3' end of the 67pro.

EXAMPLE 7

Promoter recovery by screening of a genomic library

Pindar genomic library

A Pindar sugarcane genomic library made in LambdaGEM®-11 (Promega) was plated with *E.coli* host strain KW251 on 14 NZY plates (150 mm) at a density of 24,000 pfu/plate. Duplicate lifts on nylon membranes (N+, Amersham), prehybridisation and hybridisation were carried out as for the screening of the cDNA library.

The probe was made from cDNA c322 (SEQ ID NO: 6) excised from the plasmid by an *Eco*RI-*Eco*O109I digest and randomly radiolabelled (Rediprime DNA labelling system, Amersham). Two genomic clones were isolated after a secondary screening: λ32A2 and λ32A6. The Lambda DNA was isolated from liquid lysates (Ausubel *et al.* 1990), restricted by different enzymes (*Pst*I, *Eco*RV, *Eco*RI, *Hin*dIII, *Nco*I, *Bst*XI and *Eco*O109I) and the fragments were analysed by Southern blotting. From both Lambda clones, a *Pst*I fragment containing the promoter region was cloned into pZErO-2 (Invitrogen). Sequences from the resulting plasmids (pZ-32A2P2 and pZ-32A6P2) are shown in FIGS. 19 and 20 (SEQ ID NO: 13 and 14).

H32-8560 genomic library

A sugarcane genomic library (Albert *et al.* 1992, *Plant Mol. Biol.* 20:663-671) made in Lambda EMBL4 (Stratagene) was plated with *E.coli* host strain Y1090r- on ten 150 mm-plates at a density of 35,000 pfu/plate. Duplicate lifts on nylon membranes (N+, Amersham), prehybridisation and hybridisation were carried out as for the screening of the cDNA library. A *Nhe*I DNA fragment from the cDNA c51 (SEQ ID NO: 2) was used to generate a radiolabelled probe (Rediprime DNA labelling system, Amersham). One clone was isolated after tertiary screening (λH51-6). The DNA fragments from an *Eco*RI and a *Pst*I restriction digest corresponding to hybridising signals from the clone λH51-6 were cloned into the pZErO™-2 cloning vector (Invitrogen). These clones were designated pZ-H51-6E (*Eco*RI) and pZ-H51-6P1 (*Pst*I). The pZ-H51-6E insert is 1052 bp long and contains the ATG start codon and 960 bp of sequence upstream of the ATG. The pZ-H51-6P1 insert is 3.3 kb long and contains 670 bp of sequence upstream of the ATG. The 240 bp *Eco*RI-*Pst*I fragment from pZ-H51-6E was used to re-probe the Southern blot of the λH51-6 restriction digests. A *Pst*I DNA fragment hybridising to this probe was cloned into pZErO (pZ-H51-6P2). The sequence of pZ-H51-6P2 corresponds to the sequence immediately upstream of the sequence from pZ-H51-6P1. The sequence of the promoter region of the λH51-6 clone is shown in FIG. 21 (SEQ ID NO: 15).

Fusion of the promoter region to a reporter gene

32A6

A primer was designed for amplification of the promoter region from pZ-32A6P2 in conjunction with the M13-Reverse primer (5' CAGGAAACAGCTATGAC 3'; SEQ ID NO:26). This primer sequence (primer 32A4) is homologous to the region surrounding the ATG and contains a *Bam*HI restriction site at its 5' end (in bold). The sequence is 5' TAGGATCCTACCATCTTGAGATGCGG 3' (SEQ ID NO:27). The PCR reaction was carried out with a proofreading DNA polymerase from the ELONGASE Enzyme Mix (Life Technologies). The PCR product was digested by *Pst*I and *Bam*HI and purified. The GUS gene and *nos* terminator were excised from plasmid

pBI101.1 (Clontech) as a *Xba*I-*Eco*RI fragment and ligated into the *Xba*I and *Eco*RI sites of the pGEM®-4Z (Promega) to form plasmid pG4-GUS-1. This plasmid was restricted by *Bam*HI and *Pst*I and the PCR product ligated into these two sites to form a translational fusion of the 32A6 promoter to the GUS reporter gene. This plasmid construct was designated p32A6G and its fusion region confirmed by sequencing with the primer Gus-2 (5' CGCTGATCAATTCCACAG 3'; SEQ ID NO:28).

32A2

The primer 32A4 was also used to amplify the promoter region of pZ-32A2P2 in conjunction with the M13-Reverse primer. The PCR reaction was carried out with a proofreading DNA polymerase from the ELONGASE Enzyme Mix (Life Technologies). The PCR product was restricted by *Bam*HI and *Pst*I, purified and ligated into the *Bam*HI and *Pst*I sites of pG4-GUS-1 to form p32A2G. The fusion region was confirmed by sequencing p32A2G with the primer Gus-2.

51

The 51 promoter region is present in two plasmids (pZ-H51-6P1 and pZ-H51-6P2). Fusion of the two parts of the promoter was carried out as follows: pZ-H51-6P1 was partially digested with *Pst*I. The full-length linear plasmid was gel purified. The far upstream promoter region in pZ-H51-6P2 was cut from the plasmid pZ-H51-6P2 as a *Pst*I fragment and ligated into the *Pst*I site of pZ-H51-6P1. The resulting plasmid was designated pZ-H51-6P1P2.

The GUS gene and *nos* terminator were excised from plasmid pBI101.3 (Clontech) as a *Xba*I-*Eco*RI fragment and ligated into the *Xba*I-*Eco*RI sites of the pGEM®-4Z (Promega) to form plasmid pG4-GUS-3. Two translational fusions of the 51 promoter were made in pG4-GUS-3. An oligonucleotide primer (51G) homologous to the region surrounding the ATG start codon was designed to contain a *Bam*HI restriction site (in bold) in its 5' end. The sequence of 51G is 5' ATGGATCCCCATTGGTGACGATCAGAAG 3' (SEQ ID NO:29). Using 51G and M13-Reverse primers, the promoter region was amplified from pZ-H51-6E with a proofreading DNA polymerase from the ELONGASE Enzyme Mix (Life

Technologies). The PCR product was then restricted with *Bam*HI and ligated into the *Bam*HI site of pG4-GUS3, generating the plasmid p51SG (959bp of promoter region to the ATG). The plasmid p51SG was restricted by *Nco*I and *Sal*I and a *Sal*I – *Nco*I fragment from the plasmid pZ-H51-6P1P2 was inserted, resulting in the plasmid p51LG which contains 2133 bp of promoter region (to the ATG).

EXAMPLE 8

Functionality of the promoter: sugarcane transformation

Particle bombardment

Plasmids (GUS constructs) were isolated by rapid alkaline extraction, dissolved in TE buffer and checked for quality (intactness and absence of genomic DNA and RNA contamination) by gel electrophoresis, quantified by spectrophotometry (Sambrook *et al.* 1989), and not linearised before precipitation onto tungsten particles. Tungsten (Bio-Rad M10) was purchased in small quantities, sterilised by washing in ethanol and sterile water, and stored in water at –20°C. Tungsten stored for several years as a dry powder may become unsuitable for gene transfer.

Precipitation reactions were performed by adding the following reagents at 4°C in the listed order to a 1.5 mL microfuge tube: 5 µL selectable marker plasmid DNA (1 mg/mL pEmuKN unless specified), 5 µL non-selected plasmid DNA (1 mg/mL unless specified), 50 µL tungsten (M10, 100 mg/mL thoroughly suspended in water), 50 µL $CaCl_2$ (2.5 M in water), 20 µL spermidine (100 mM free base in water). The preparation was mixed immediately after addition of each reagent, with minimal delay between addition of $CaCl_2$ and spermidine. The tungsten was then allowed to settle for 5 min on ice, before removal of 100 µL of supernatant and resuspension of the tungsten by running the tube base across a tube rack. Suspensions were used within 15 min, at a load of 4 µL/bombardment, with resuspension of particles immediately before removal of each aliquot. Assuming the entire DNA is precipitated during the reaction, this is equivalent to 1.3 µg DNA/bombardment, on 667 µg tungsten/bombardment.

Particles were accelerated by direct entrainment in a helium gas pulse, through the constriction of a syringe filter holder into target tissues in a vacuum chamber, as described previously (Bower *et al.* 1996, *supra*).

Target tissues, selection and plant regeneration

Embryogenic callus cultures of sugarcane cultivar Q117 were initiated and bombarded, with osmotic treatment, followed by escape-free selection for resistance to 45 mg/L G418, and regeneration to plants as described previously (Bower *et al.* 1996, *supra*).

Potted plants were grown at 28 ± 2°C in a containment glasshouse without artificial lighting.

EXAMPLE 9

Analysis of transgenic sugarcane plants

Southern analysis

Regenerated plants were confirmed transgenic by Southern analysis, using the GUS coding region as a probe.

Histochemical GUS analysis

Histochemical detection of GUS activity was done using the substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc; Jefferson 1987, *Plant Molecular Biology Reporter* 5:387-405). Sugarcane tissues (leaf segments, roots and stem slices) were immersed in assay buffer (0.05% X-Gluc, 50 mM sodium phosphate pH 7.0), vacuum infiltrated twice for 10min and incubated for 16hrs at 37°C. Following this incubation, green tissues were destained with ethanol and blue-stained tissues identified. FIG. 22 shows the GUS activity in the stem of three transgenic plants transformed with plasmid p67G. This is the first demonstrated recovery of a functional promoter from sugarcane and retention of the promoter activity and specificity when re-introduced into sugarcane, which has so far shown a marked propensity to silence transgenes.

The present invention has been described in terms of particular embodiments found or proposed by the present inventors to comprise preferred modes for the practice of the invention. Those of skill in the art will appreciate that, in light of the present disclosure, numerous modifications and changes may be made in the particular embodiments exemplified without departing from the scope of the invention.

CLAIMS

1. An isolated nucleic acid comprising a nucleotide sequence which corresponds to a promoter region of a transcribable DNA sequence that is hybridizable to a probe or primer derivable from a polynucleotide sequence selected from the group consisting of: -

(a) the polynucleotide sequence set forth in FIG. 14 [SEQ ID NO: 1];

(b) the polynucleotide sequence set forth in FIG. 15 under designator c51 [SEQ ID NO: 2];

(c) the polynucleotide sequence set forth in FIG. 15 under designator c511 [SEQ ID NO: 3];

(d) the polynucleotide sequence set forth in FIG. 15 under designator c512 [SEQ ID NO: 4];

(e) the polynucleotide sequence set forth in FIG. 16 [SEQ ID NO: 5];

(f) the polynucleotide sequence set forth in FIG. 17 [SEQ ID NO: 6];

(g) the polynucleotide sequence set forth in FIG. 23 [SEQ ID NO: 7];

(h) the polynucleotide sequence set forth in FIG. 24 [SEQ ID NO: 8];

(i) the polynucleotide sequence set forth in FIG. 26 [SEQ ID NO: 9];

(j) the polynucleotide sequence set forth in FIG. 29 [SEQ ID NO: 10]; and

(k) the polynucleotide sequence set forth in FIG. 31 [SEQ ID NO: 11].

2. An isolated nucleic acid comprising a nucleotide sequence which corresponds to a promoter region of a transcribable DNA sequence selected from the group consisting of: -

(a) the polynucleotide sequence set forth in FIG. 14 [SEQ ID NO: 1];

(b) the polynucleotide sequence set forth in FIG. 15 under designator c51 [SEQ ID NO: 2];

(c) the polynucleotide sequence set forth in FIG. 15 under designator c511 [SEQ ID NO: 3];

(d) the polynucleotide sequence set forth in FIG. 15 under designator c512 [SEQ ID NO: 4];

(e) the polynucleotide sequence set forth in FIG. 16 [SEQ ID NO: 5];

(f) the polynucleotide sequence set forth in FIG. 17 [SEQ ID NO: 6];

(g) the polynucleotide sequence set forth in FIG. 23 [SEQ ID NO: 7];

(h) the polynucleotide sequence set forth in FIG. 24 [SEQ ID NO: 8];

(i) the polynucleotide sequence set forth in FIG. 26 [SEQ ID NO: 9];

(j) the polynucleotide sequence set forth in FIG. 29 [SEQ ID NO: 10]; and

(k) the polynucleotide sequence set forth in FIG. 31 [SEQ ID NO: 11].

3. The isolated nucleic acid of Claim 1, wherein the nucleotide sequence is of a length in the range 100 bp to 4 kb.

4. The isolated nucleic acid of Claim 3, wherein the nucleotide sequence is of a length in the range 1 kb to 4kb.

5. The isolated nucleic acid of Claim 1, wherein the isolated nucleic acid is capable of directing transcription in many or all tissues of a plant.

6. The isolated nucleic acid of Claim 1, wherein the isolated nucleic acid is capable of directing transcription preferentially in stem tissue of a plant.

7. The isolated nucleic acid of Claim 1, wherein the isolated nucleic acid is capable of directing transcription preferantially in meristem tissue of a plant.

8. The isolated nucleic acid of any one of Claims 5-7, wherein the plant is a monocotyledonous plant.

9. The isolated nucleic acid of Claim 8, wherein the monocotyledonous plant is sugarcane.

10. An isolated nucleic acid which comprises a nucleotide sequence selected from the group consisting of: SEQ ID NO: 12 , SEQ ID NO:13, SEQ ID NO: 14 and SEQ ID NO:15.

11. An isolated nucleic acid which comprises a biologically-active fragment of any one of SEQ ID NOS: 12, 13, 14 or 15.

12. An isolated nucleic acid which has at least 60% sequence identity with any one of SEQ ID NOS: 12, 13, 14 or 15.

13. An isolated nucleic acid which is capable of hybridising to any one SEQ ID NO: 12, 13, 14 or 15 under at least low stringency conditions.

14. A chimeric gene comprising the isolated nucleic acid of any one of Claims 1, 2, 10, 11, 12 or 13 operably linked to a heterologous nucleic acid.

15. The chimeric gene of Claim 14, characterized in that said isolated nucleic

acid is capable of directing transcription of the heterologous nucleic acid in many or all tissues of a plant.

16. The chimeric gene of Claim 14, characterized in that said isolated nucleic acid is capable of directing transcription of the heterologous nucleic acid preferentially in stem tissue of a plant.

17. The chimeric gene of Claim 14, characterized in that said isolated nucleic acid is capable of directing transcription of the heterologous nucleic acid preferentially in meristem tissue of a plant.

18. An expression vector comprising the isolated nucleic acid of any one of Claims 1, 2, 10, 11, 12 or 13.

19. The expression vector of Claim 14, wherein the isolated nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 12, 13, 14 or 15.

20. The expression vector of Claim 18 or 19, further comprising a heterologous nucleic acid operably linked to said isolated nucleic acid.

21. The expression vector of Claim 20, wherein the heterologous nucleic acid encodes a molecule which inhibits the expression of a target gene.

22. The expression vector of Claim 21, wherein the molecule is an antisense RNA.

23. The expression vector of Claim 21, wherein the molecule is a ribozyme.

24. The expression vector of Claim 21, wherein the molecule is capable of inhibiting expression of the target gene by co-suppression.

25. The expression vector of Claim 20, wherein the heterologous nucleic acid encodes a polypeptide.

26. A method of transforming a plant cell or tissue, comprising the step of introducing into said plant cell or tissue the isolated nucleic acid of any one of Claims 1, 2, 10, 11, 12 or 13.

27. A transformed plant cell or tissue comprising the isolated nucleic acid of any one of Claims 1, 2, 10, 11, 12 or 13.

28. A transgenic plant comprising the isolated nucleic acid of any one of Claims 1, 2, 10, 11, 12 or 13.

29. The transgenic plant according to Claim 28, wherein the transgenic plant has

an altered phenotype.

30. A transgenic monocotyledonous plant according to Claim 29.

31. A transgenic sugarcane plant according to Claim 30.

32. A progeny plant produced from the transgenic plant of Claim 28

33. A cell, tissue or seed obtained from the plant of any one of Claims 28-31.

34. Plasmid pG4-67pro deposited with AGAL on August 30, 1999 under accession number NM99/05995.

35. Plasmid pZ-32A2P2 deposited with AGAL on August 30, 1999 under accession number NM99/05994.

36. Plasmid pZ-32A6P2 deposited with AGAL on August 30, 1999 under accession number NM99/05993.

37. Plasmid pZ-H51-6P1P2 deposited with AGAL on August 30, 1999 under accession number NM99/05992.

38. An isolated nucleic acid comprising a polynucleotide sequence selected from the group consisting of: -

(a) the polynucleotide sequence set forth in FIG. 14 [SEQ ID NO: 1];

(b) the polynucleotide sequence set forth in FIG. 15 under designator c51 [SEQ ID NO: 2];

(c) the polynucleotide sequence set forth in FIG. 15 under designator c511 [SEQ ID NO: 3];

(d) the polynucleotide sequence set forth in FIG. 15 under designator c512 [SEQ ID NO: 4];

(e) the polynucleotide sequence set forth in FIG. 16 [SEQ ID NO: 5];

(f) the polynucleotide sequence set forth in FIG. 17 [SEQ ID NO: 6];

(g) the polynucleotide sequence set forth in FIG. 23 [SEQ ID NO: 7];

(h) the polynucleotide sequence set forth in FIG. 24 [SEQ ID NO: 8];

(i) the polynucleotide sequence set forth in FIG. 26 [SEQ ID NO: 9];

(j) the polynucleotide sequence set forth in FIG. 29 [SEQ ID NO: 10]; and

(k) the polynucleotide sequence set forth in FIG. 31 [SEQ ID NO: 11].

39. A polypeptide encoded by an open reading frame of the isolated nucleic acid

according to Claim 38.

40. The polypeptide of Claim 39, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of:

(i) the amino acid sequence set forth in FIG. 14 [SEQ ID NO:16];

(ii) the amino acid sequence set forth in FIG. 15 [SEQ ID NO:17];

(iii) the amino acid sequence set forth in FIG. 26 [SEQ ID NO:18];

(iv) the amino acid sequence set forth in FIG. 29 [SEQ ID NO:19];

(v) the amino acid sequence set forth in FIG. 31 [SEQ ID NO:20];

(vi) the amino acid sequence set forth in FIG. 24 [SEQ ID NO:21];

(vii) the amino acid sequence set forth in FIG. 16 [SEQ ID NO:22]; and

(viii) the amino acid sequence set forth in FIG. 17 [SEQ ID NO:23].

DATED this first day of September, 2000

THE UNIVERSITY OF QUEENSLAND

By their Patent Attorneys

FISHER ADAMS KELLY

Fig. 1

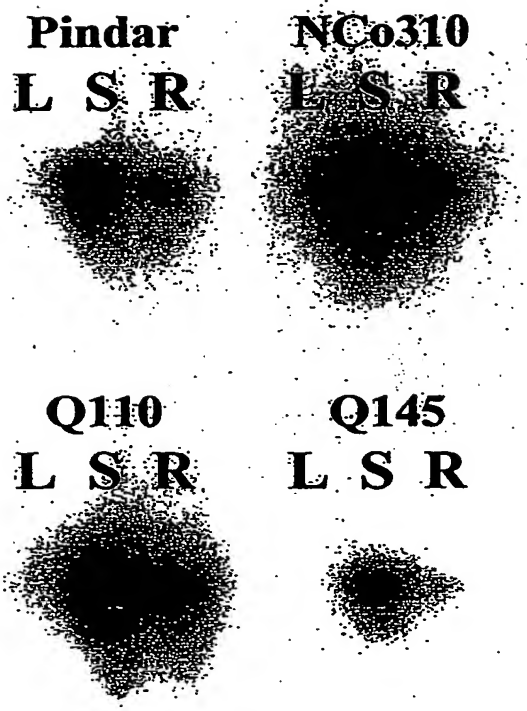


Fig. 2

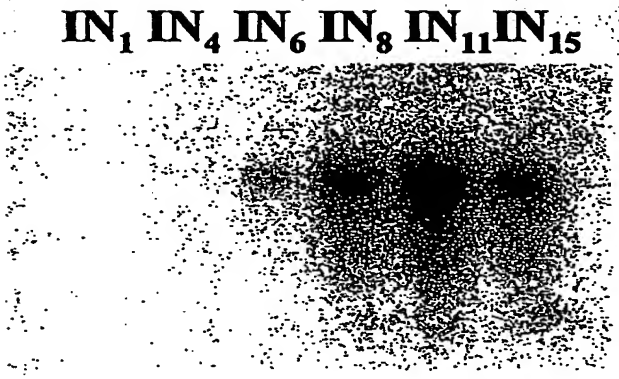


Fig. 3

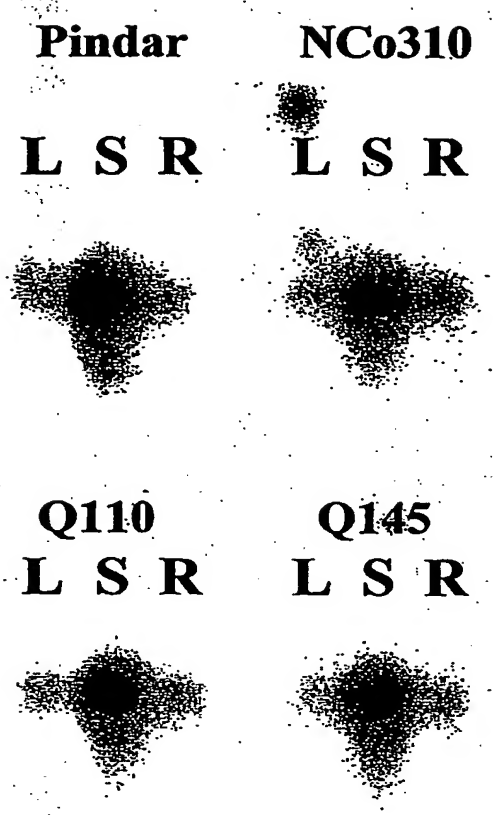
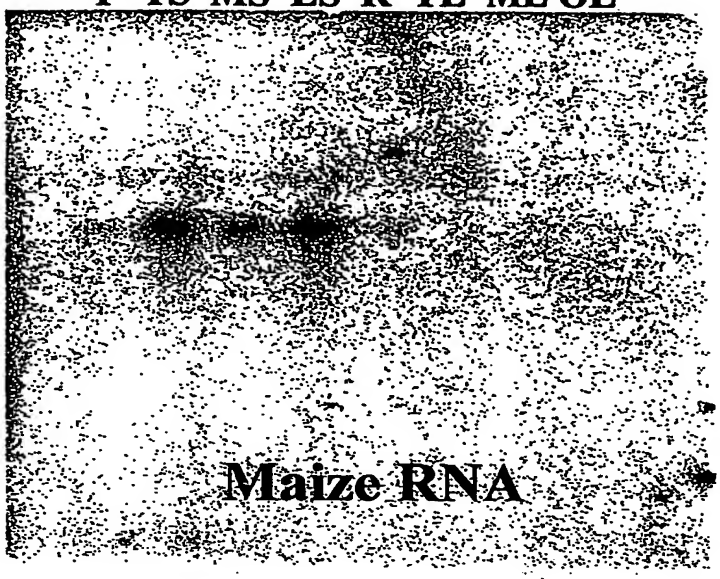


Fig. 4

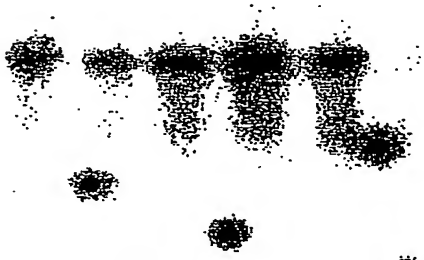
$IN_6IN_{11}\ N_c\ N\ M\ R\ YL\ ML\ OL$

Fig. 5

F TS MS LS R YL ML OL



P TS MS LS R YL ML OL

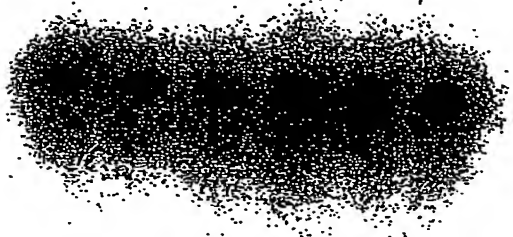


Sorghum RNA

Fig. 6

Sugarcane RNA

IN_1 IN_4 IN_6 IN_8 IN_{11} IN_{15}



Maize RNA

IN_1 IN_3 IN_5 IN_7 IN_9 IN_{11}

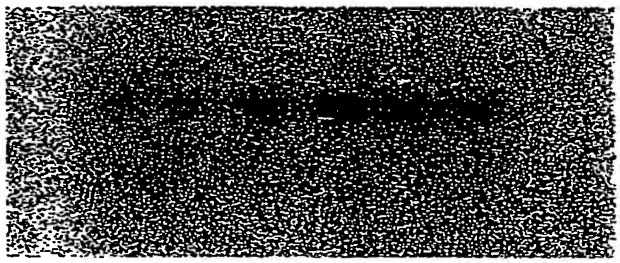


Fig. 7

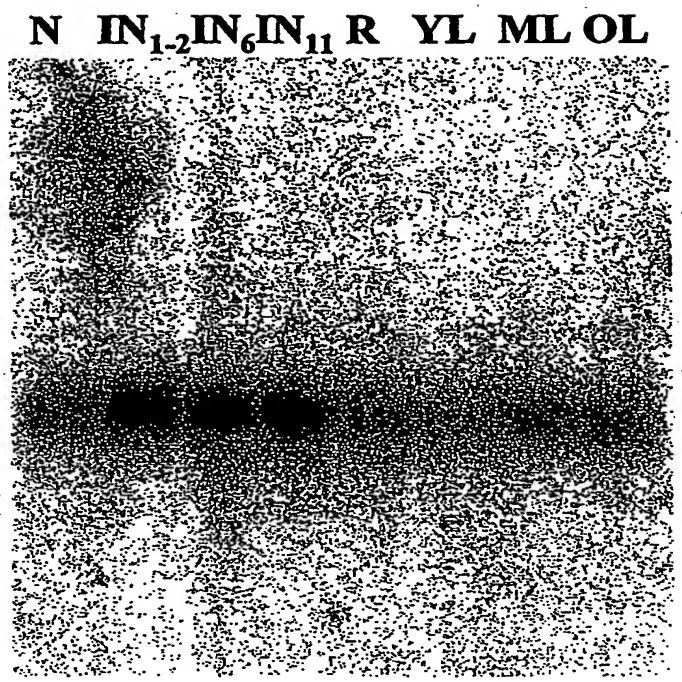


Fig. 8

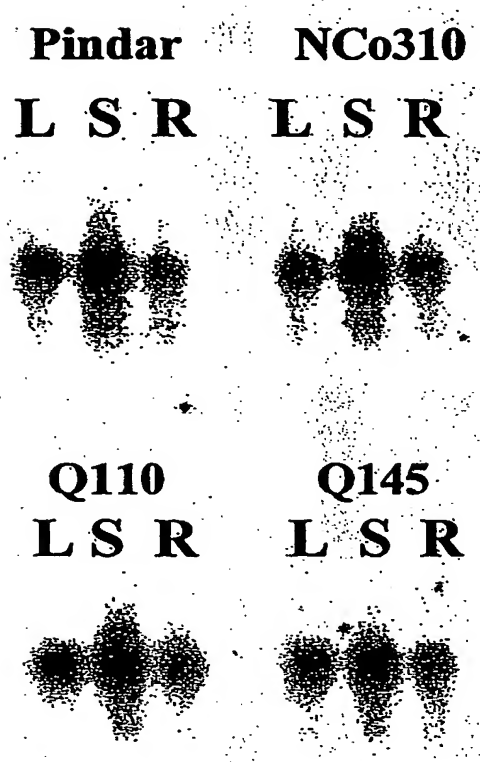


Fig. 9

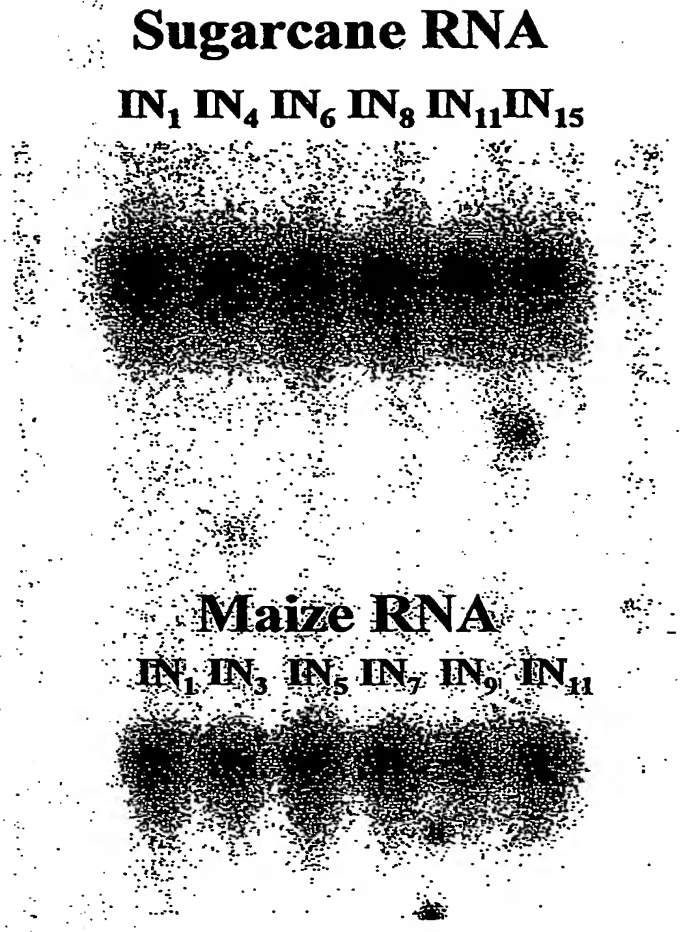
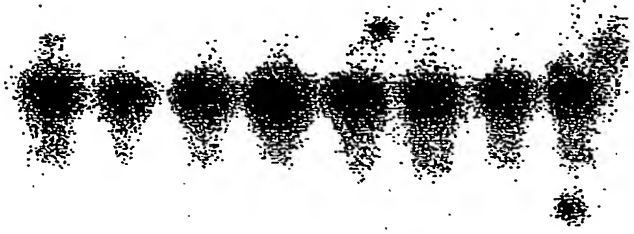

Sugarcane RNA
IN1 IN4 IN6 IN8 IN11IN15
Maize RNA
IN1 IN3 IN5 IN7 IN9 IN11

Fig. 10

Sorghum RNA

P TS MS LS R YL ML OL



Rice RNA

S_1 S_2 R L_1 L_2 L_3

Fig. 11

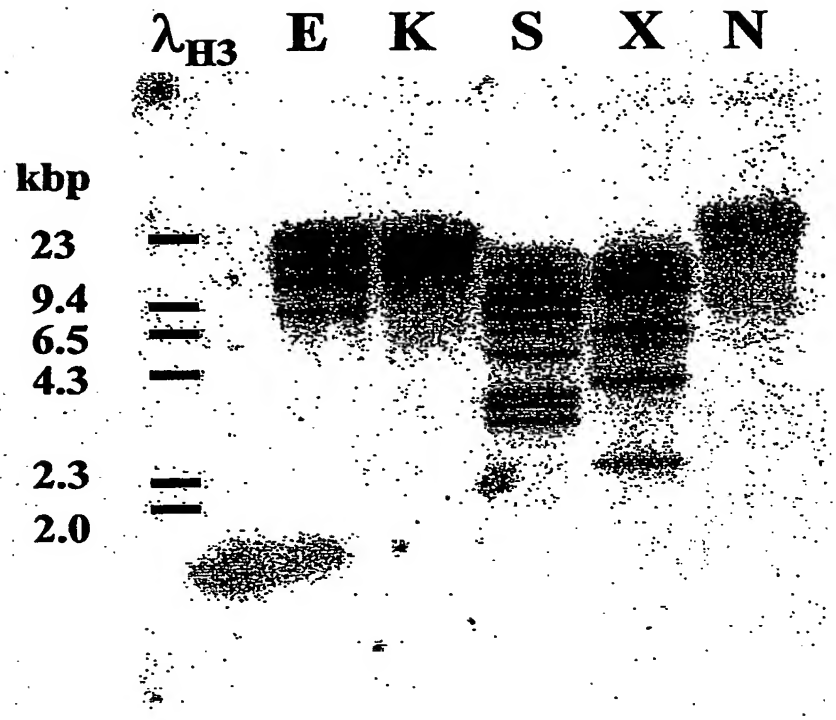


Fig. 12

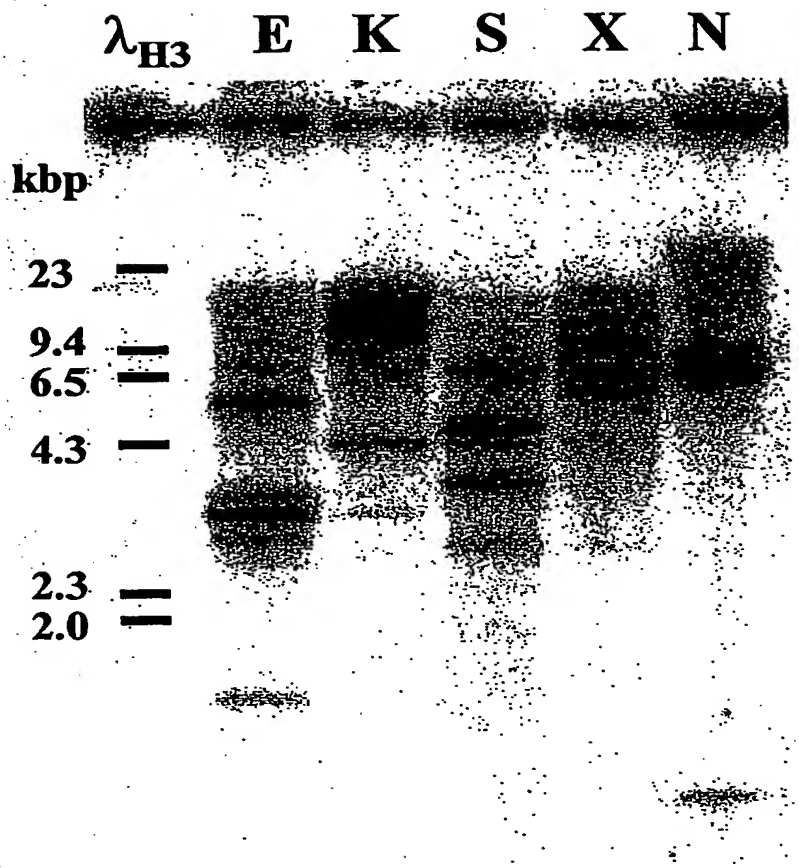


Fig. 13

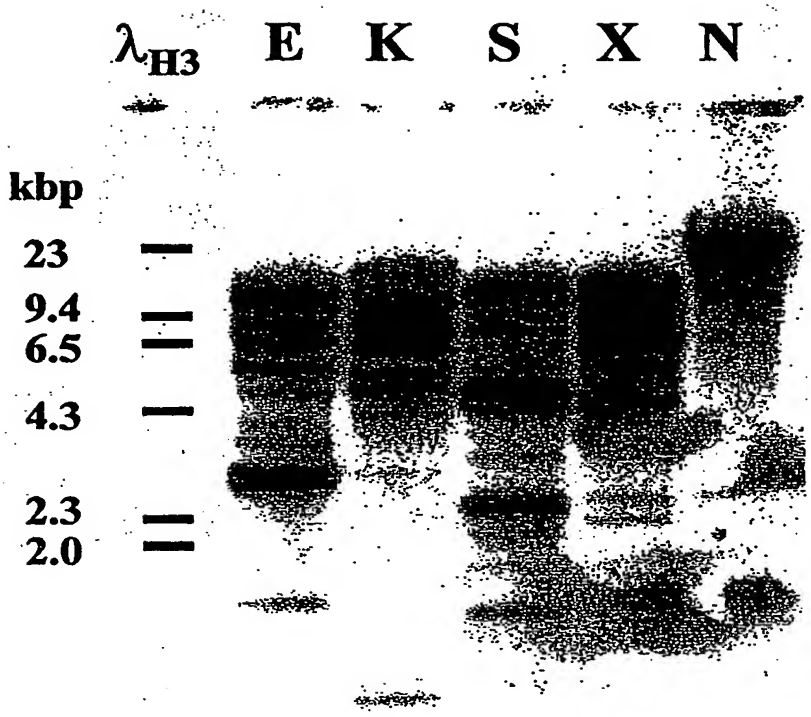


Fig.14

```
1   TGAAATAAAC GGTAGCTGCC ATAACTAGTA CAATGGCCAA TCTGAGAAGC   50
                                    M  A  N  L  R  S
51  GTCCTAGCTG TGAGCCTAGC CGTGGCACTT TTCGCAGTTG CTCCTGCATC  100
    V  L  A  V  S  L  A  V  A  L  F  A  V  A  P  A  S
101 GTTCGCACTG GATGAGAAAG AGTTGCACCT GAGTTTGTAC TTAAACCAGA  150
     F  A  L  D  E  K  E  L  H  L  S  L  Y  L  N  Q  T
151 CATACAGCGG AAACGGCCTT AACCAGGCGG TGGTGGTCGA ACCAGGCCTA  200
      Y  S  G  N  G  L  N  Q  A  V  V  V  E  P  G  L
201 CCTGGGGAGT TCGGCAACAT CGCCGTCCAG GACTGGCCTG TGACCAATGG  250
    P  G  E  F  G  N  I  A  V  Q  D  W  P  V  T  N  G
251 GGAAGGTAGC GACGCAACCG TCGTTGGACG TGCACAGGGC ATCCAGTTCA  300
     E  G  S  D  A  T  V  V  G  R  A  Q  G  I  Q  F  K
301 AACCAAGCCA GAGGAACGAC CAAGCCTGGT ATACCACCTT GACCATAGTG  350
      P  S  Q  R  N  D  Q  A  W  Y  T  T  L  T  I  V
351 TTCGAGAACA CGAGCCTCAA GGGATCCACG CTTCAGATGA TGGGTTACAT  400
    F  E  N  T  S  L  K  G  S  T  L  Q  M  M  G  Y  I
401 CCCACAAGAT GGTCAGTGGA GCATTTTTGG AGGAACTGGA CAACTTACGA  450
     P  Q  D  G  Q  W  S  I  F  G  G  T  G  Q  L  T  M
451 TGGCACGCGG TGTTGTGAAC CACAAGGTTG TGCGCCAAAC CAATGGCGGG  500
      A  R  G  V  V  N  H  K  V  V  R  Q  T  N  G  G
501 AGGATGTATA AGATCAACAT ACATGCCTTC TATACCCCCC TGGGCGCTTC  550
    R  M  Y  K  I  N  I  H  A  F  Y  T  P  L  G  A  S
551 TAGCAACTGT GGGATTAACC TTAGGCGCTT GGACTTCGAC GCTTGATCGA  600
     S  N  C  G  I  N  L  R  R  L  D  F  D  A  *
601 CTAGCGCGGA CTACAACAGG AGGACCGTGT TCTTCGTCGA CGCTTAATGC  650
651 ATGGAAACTT CCCACGGGGG ACCGTGTTCT TCATGGACGC TAGACCAACC  700
701 ATAATTTCTT TTCCGTTTGT ACTGTCAACA AATATAAATA TGTAAAGCAT  750
751 AAATCCGAAC TGTATTCGGA GTTGCAGGTC GTCGTTGCCC CTGCCTTATG  800
801 GCCTGCACTG TACATGTACA TGTTTCTGTC AAGTTCTGCG AGTATTTTAA  850
851 GTAATAAATA AAGTGGTTGG TTTCACGGTT TAAAAAAAAA AAAAAAAAAA  900
901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA  950
951 AAAAAAAAAA AAAAAAAAAA AAAAAA  976
```

Fig.15

```
          1                                                   50
 c51      TCGAATGCTC GATCGATCCC ACTCTCAGCT GATCGCTCAC TCTTGCAGCT
c511      .....TGCTC GATCGATCCC ACTCTCAGCT GATCGCTCAC TCTTGCAGCT
c512      ......TCTC GATCGATCCC ACTCTCAGCT GATCGCTCAC TCTTGCAGCT

          51                                                  100
 c51      CGATCAGTCT TAGCTCTAGC CTCTAGCTAG CCAACTAGCC GCTCCTTCGT
c511      CGATCAGTCT TAGCTCTAGC CTCTAGCTAG CCAACTAGCC ACTCCTTCGT
c512      CGATCAGTCT TAGCTCTAGC CTCTAGCTAG CCAACTAGCC ACTCCTTCGT

          101                                                 150
 c51      GTAGCCATCA GCCTTCTGAT CGTCACCAAT GGCCACTGCC GAGGTCCAGA
c511      GTAGCCATCA GCCTTCTGAT CGTCACCAAT GGCCACTGCC GAGGTCCAGA
c512      GTAGCCATCA GCCTTCTGAT CGTCACCAAT GGCCACTGCC GAGGTCCAGA
aa                                      M   A T A   E V Q T

          151                                                 200
 c51      CCCCGACCGT CGTGGCGACC GAGGAGGCGC CCGTGGTGGA GACGCCGGCG
c511      CCCCGACCGT CGTGGCGACC GAGGAGGCGC CCGTGGTGGA GACGCCGGCG
c512      CCCCGACCGT CGTGGCGACC GAGGAGGCGC CCGTGGTGGA GACGCCGGCG
aa          P T V  V A T  E E A P  V V E  T P A

          201                                                 250
 c51      CCGGCCGTCG TGCCCGAGGA GGCTGCCCCC GCCCCCGCCG AGGCTGAGAC
c511      CCGGCCGTCG TGCCCGAGGA GGCTGCCCCC GCCCCCGCCG AGGCTGAGCC
c512      CCGGCCGTCG TGCCCGAGGA GGCTGCCCCC GCCCCCGCCG AGGCTGAGCC
aa        P A V V  P E E  A A P  A P A E  A E P

          251                                                 300
 c51      GGCGGCCGTG CCCGAGGAGG CTGCCCCCGC CGAGGCCAAG GTGGAGGAGC
c511      GGCGGCCGTG CCCGAGGAGG CTGCCCCCGC CGAGGCCAAG GTGGAGGAGC
c512      GGCGGCCGTG CCCGAGGAGG CTGCCCCCGC CGAGGCCAAG GTGGAGGAGC
aa         A A V  P E E A  A P A  E A K  V E E P

          301                                                 350
 c51      CTGCCGCCCC GGCGGAGCCT GAGCCTGCCG CCGCTGAGCC CGAGGCCGAG
c511      CTGCCGCCCC GGCGGAGCCT GAACCTGCCG CCGCTGAGCC CGAGGCCGAG
c512      CTGCCGCCCC GGCGGAGCCT GAGCCTGCCG CCGCTGAGCC CGAGGCCGAG
aa          A A P  A E P  E P A A  A E P  E A E

          351                                                 400
 c51      CCTGCCGCCG CGGAGCCGGA GGCGGCCCCT GCAGCCGCGG CGGAGGAAGA
c511      CCTGCCGCCG CGGAGCCGGA GGCGGCCCCT GCAGCCGCGG CGGAGGAAGA
c512      CCTGCCGCCG CGGAGCCGGA GGCGGCCCCT GCGGCCGCGG CGGAGGAAGA
aa        P A A A  E P E  A A P  A A A A  E E E

          401                                                 450
 c51      GGCGCCAAAG GAGGCGGAGC CGGCGGCGGT TGAGGAGGTG AAGGAGGAGG
c511      GGCGCCAAAG GAGGCGGAGC CGGCGGCGGT TGAGTAGGTG AAGGAGGAGG
c512      GGTGCCAAAG GAGGCGGAGC CGGCGGCGGT TGAGGAGGTG AAGGAGGAGG
aa         A(v)P K  E A E P  A A V  E E V  K E E E

          451                                                 500
 c51      AGGCGGCGGC GCCCGCTGCC GAGACAGAGC CGGCGGCCGC CGAGCCCGAG
c511      AGGCGGCGGC GCCCGCTGCC GAGACAGAGC CGGCGGCCGC CGAGCCCGAG
c512      AGGCGGCGGC GCCCGCTGCC GAGACAGAGC CGGCGGCCGC CGAGCCCGAG
aa          A A A  P A A  E T E P  A A A  E P E
```

Fig.15 cont'd

```
      501                                                        550
 c51  GTTGCTGCTC CTGCTGCCTC CGAGGAGCCC ACCGCGGCCG AGCCGGCCGC
c511  GCTGCTGCTC CTGCTTCCTC CGAGGAACCC GCCGCGGCCG AGCCGGCCGC
c512  GCTGCTGCTC CTGCTGCCTC CGAGGAGCCC ACCGCGGCCG AGCCGGCCGC
aa    V(A)A  A  P   A  A  S   E  E  P   T  A  A  E   P  A  A

      551                                                        600
 c51  CGCGGAGCCC GAGAAGGCCA GCGAGTGAGG CCGTGCGCGC GCGCAGCGGC
c511  CGCGGAGCCC GAGAAGGCCA GCGAGTGAGG CCGT..GCGC GCGCAGCGGC
c512  CGCGGAGCCC GAGAAGGCCA GCGAGTGAGG CCGTGCGCGC GCGCAGCGGC
aa     A  E  P   E  K  A  S   E  *

      601                                                        650
 c51  GGCGGCCAGG GGATCGGAGT GAGATGGCTC ATCGTGCGCG CGTGACGTAG
c511  GGCGGCCAGG GGATCGGAGT GAGGTGGCTC ATCGTGCGCG CGTGACGTAG
c512  GGCGGCCAGG GGATCGGAGT GAGATGGCTC ATCGTGCGCG CGTGACGTAG

      651                                                        700
 c51  CGGCGTGGCA GTCGCGTGCG CGCGTACATG GCACGGTCTT TTGCTGTTCA
c511  CGGCGTGGCA GTCGCGTTTG CGCGTACATG GCACGGTCTT TTGCTGTTCA
c512  CGGCGTGGCA GTCGCGTGCG CGCGTACATG GCACGGTCTT TTGCTGTTCA

      701                                                        750
 c51  GTACGCTACG GTTACTGGCG TTTC..TGTG TGTGTGTCGT AGCTCGGTAC
c511  GTACGCTACG GTTACTGGCG TTTCTGTGTG TGTGTGTCGT AGCTCGGTAC
c512  GTACGCTACG GTTACTGGCG TTTC..TGTG TGTGTGTCGT AGCTCGGTAC

      751                                                        800
 c51  GTAGCTAGCA GTGGCTGTCG CGTGGACTGT CCGGTGTGGA CTGCGAGTGA
c511  GTAGCTAGCA GTGGCTGTCG CGTGGACTGT CCGGTGAGGA CTGCGAGTGA
c512  GTAGCTAGCA GTGGCTGTCG CGTGGACTGT CCGGTGTGGA CTGCGAGTGA

      801                                                        850
 c51  CGTGGGCTAA ATAAAGTGTG CGTCGTGTGG TGG...TGGT GGTGGGGCCC
c511  CGTGGGCTAA ATAAAGTGTG CGTGGTGGTG GGGCCCGGTG TCAGTGGCGC
c512  CGTGGGCTAA ATAAAGTGTG CGTCGTGTGG TGG...TGGT GGTGGGGCCC

      851                                                        900
 c51  GGTGCCCCGT GCCCTGCCGC TGCAGCTAGT CTGTCTTTT. .GTGTTTGTT
c511  GGTGCCCCAT GCCCTGCCGC TGCAGCTTGT CTGTCTTTT. GTGTCGTGTT
c512  GGTGCCCCGT GCCCTGCCGC TGCAGCCTGT CTGTCTTTTT GTGTTGTGTT

      901                                                        950
 c51  GGTTCCTGTT CTCTGTCGAT CTATATTTAT CCTCATGTTT GTATGGCAGT
c511  GGTTCCTGTT CTCTGTCGAT CTATATTTAT CCTCATGTTT GTATGGCAGT
c512  GGTTCCTGTT CTCTGTCGAT CTATATTTAT CCTCCA31... ..........

      951                              979
 c51  TTGGGTATGG GTATCA49... .........
c511  TTGGGTATGG GTATCAAACA TGTTTGCGA30
c512  .......... .......... .........
```

Fig.16

```
  1  TGCCAGGCTC ACCGAGGGGT GCTCCTTCCG TCGCAAGGGC GATTAGATCC   50
       A  R  L   T  E  G  C   S  F  R   R  K  G   D  *
 51  TGCTGTCTTC ATGGGAGAAG AGAGAAATTT GCTGTTTTAC TACCTTCCCA  100
101  TGATATGTAC TCGTTGAGGA TTTTGTTGAT TATTATGGCT GTTTAGCGTG  150
151  CCCTGGCAAT GCTTTTGTAA ACGTGCACTT TGCTTGAGCT TAGTGACATC  200
201  TACTAAGGTG CTGTTTGGTT TTGCTAAGGG AGTGGCAATG GTAATGAAAT  250
251  CAGTTGCTGG CGTTAATTGT TGGGGTTTTA AAAAAAAAAA AAAAAAAAAA  300
301  AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA  350
351  AAAAAAAAAA AAAAAAAAAA AAAAAAAA  378
```

Fig.17

```
  1  GCCGCCGCTA CGCGCTCGCC TTTGCTCCCT CTGTTTCCCT TCCTCCATAC   50
 51  AGGCGCAGGC CAGGAACGCA CAAGGCGACC GCATCTCCAA GATGGTGCTG   100
                                                  M  V  L
101  CAGAACGACA TCGACTTGCT CAACCCGCCG GCAGAGCTTG AGAAGCTAAA   150
     Q  N  D  I   D  L  L  N  P  P   A  E  L  E   K  L  K
151  GCACAAGAAG AAGCGCCTCG TCCAGTCCCC CAACTCCTTC TTCATGGATG   200
      H  K  K   K  R  L  V   Q  S  P   N  S  F   F  M  D  V
201  TCAAGTGCCA GGGCTGTTTC AGCATAACCA CTGTGTTCAG CCACTCCCAG   250
       K  C  Q   G  C  F   S  I  T  T   V  F  S   H  S  Q
251  ACTGTGGTTG TGTGCCCAGG CTGCCAAACT GTTCTCTGCC AACCTACTGG   300
     T  V  V  V   C  P  G   C  Q  T   V  L  C  Q   P  T  G
301  TGGGAAGGCC AGGCTCACCG AGGGGTGCTC CTTCCGTCGC AAGGGCGATT   350
      G  K  A   R  L  T  E   G  C  S   F  R  R   K  G  D  *
351  AGATCCTGCT GCCTTTTAAT GGGAGAAGAG AGAAATTTGC TGTTTTACAA   400
401  CCTTCCCATG ATATGTACTC GTTGAGGATT TTGTTAATTA TTATGGCTGT   450
451  TTAGCTTGCC CTGTCAATGC TTTTGTAAAC GTGCACTTTG CTTGGTGCTG   500
501  TTTGGTTTTG CCAAGGGATT GGCAATGGTA GTGAAATCAG TTGCTGACGT   550
551  TAAAAAAAAA AAAAAAA  567
```

Fig.18

```
   1  GAGCTCTCAA TGGGAGGTGC TCGAAGACAT ATTACCCAAG TGTATGGCAA   50
  51  GATGTTTAGC TAGTAACTGA CTGATAGTGT AAACGATCTC CAATGGGGCA  100
 101  AGACATATTA CCTAAGGCCA GGCTGGTTTT TGCAAGTTCG AGTAGGATAT  150
 151  AGAGATTCTC GTGCGAGTTG TAAACGATCT CCAATGGGGC AAGACATCCT  200
 201  ACCCTATATA TAGTGAAGGG GCAGTAGCTG ATTGAGAATC AATCAATCAA  250
 251  GCACAATATA ATTTATTAAT TTTTTATTCA AACCCAATTT TTTTCCTTTT  300
 301  CCAACCCTAA TTATAGTTCT CCTTTTGCCT CTAGGACAAA TTGACGTGTT  350
 351  CCGGGTATCC TGCTGAATCA AGAACAACCC TAGGTGCACC TGTCCCGATA  400
 401  GAGTCCCACC TGGGTAGGCA TTCATAGGGA TTCGGGTATT TCCTGCAAAA  450
 451  AAGCGATTAA GCTGGCTTCT AAAACTGGCT AGGCCGGATT CTGTGGCCTT  500
 501  CACTACCAGG TGATTTTCAT GTGATCCGTG CATTCTAGCA CTTTGCTATG  550
 551  TAACCCAAAC TGAAGTCGAC AACTATAAAT ATGCTACTTG CAGGATGTTA  600
 601  TCACGACACA ACTCCCAGTC TACGAAGCCT AAGTTTAGTT TTGCTCGGAG  650
 651  ACAAGCAATT GTGGCCAGTC ACTATAGCTA CGTCAGAGGG TAGTGGGAGC  700
 701  AGTTGCGTCG TTGGATTGAA AACAGGTGGA TCGTATCAGA TATTATGCAT  750
 751  TCACATGAAC AGTAAATGTG GTACAGTACT TCGCAAACAA TAAAATCTGT  800
 801  CACAATTTAT TAGTGCACTC CTGTGACGTA AATGCTTCTA CGTCAGAGGA  850
 851  TTTGAGTCCG AGGGCTGCTG CACCCATCAC TAATGACGGT CTTTACCCAT  900
 901  CATCATGGAC CATTGTTCAC ATCCATGCTA TCACTGTCGT CCTGTCCATG  950
 951  CACTGCAGCC CTCTATAAAT ACTGGCACCC CTCCCCCGTT CACAGATCAC 1000
1001  ACCACACAAG CAAGAAATAA ACGGTAGCTG CCATAACTAG TACAATG   1047
```

Fig.19

```
   1  CTGCAGAACG ACATCGACTT GCTCAACCCG CCGGCAGAGC TTGAGAAGCT   50
  51  AAAGCACAAG AAGAAGCGCC TCGTCCAGTC CCCCAACTCC TTCTTCATGG  100
 101  TATACCTCTT CATGCCCTTT TCTAATTCTG CGTGATGGGG TGTTGCGAAT  150
 151  CATGCCCTAA AGAGACCCAA CGAATGTATC ATTGGAATTT AGTTTTGGCA  200
 201  AAATAAAGAC CTGCACCTAA TAGGGTGATT TGGTTTGGAA GAGAGTCCAG  250
 251  AAATTAGATT TAGTATCGTT AACTACATCA AAAATTACTT CCAACAAAGC  300
 301  AACAAGACAG CTTAACTGCT CTTGTTTGCC ATGTAACATA CATGTAGTCA  350
 351  AGCTGGATGC TATAGGTTAT GCATTTTTCA TCAAATAAGT TGTGTTTATA  400
 401  AGAGAGCTAA AATGTGAAAA CAGAAATGAA AATGAATCTT AAAGTTACAA  450
 451  TGTATCGACT TATATTCATA TTAATTAATA TTAATATTAG TCGTGTCTGC  500
 501  TCTTGTGCCT GATCGGAGAC TGACTCTAAT CATGTTATTT TTCGTTTTTT  550
 551  TTCTCCTTTG TGTGCAGGAT GTCAAGTGCC AGGGCTGTTT CAGCATGTAA  600
 601  GTGTTCATGA TTCTTGCTCC TTTTTTGTTT TTAGTTACTA CAGTTGCAAT  650
 651  TGATATAGAC GCCGTGTCAT TCCATTTATC CATTGCTGAT TGCTGATGTG  700
 701  ATCTCCTAAT CTTGTGTTAC TGGTTTTCCA TTGACAGTTT AGCATGATCA  750
 751  GTTTACTAGT AGGGTTGCGA TTACTACTAG CTAGTTAAAA TATGAAAATT  800
 801  CTTGGTTTAG TTATAAGGTT ATATATTGAT TTCGTAAAAT TTCATACCCC  850
 851  CTCACTTCCC GAATATGTAG GACAACAAAA TTGTTTATAG ATGGAGATGG  900
 901  AAACTGTTAA TGTTTGACTC CTGTTTTTGT TTCTTTCTCA CCTGACCAGT  950
 951  ACAAGTACAA TTGTTCTGTT TAAATGTGGG TTAATTTGGA TTCAACAACA 1000
1001  ACAACAACAA CAACATAGCC TTTTGTCCCA AGCAAGTTGG GGTAGGCTAG 1050
1051  AGATGAAACC CAAAAGAGAC GAGAAACAGG GAGACACAAC GTTACACCTC 1100
1101  GCTGCTTTAT TTGGATTACG CTTGTATTTC TTTTGATAAT TGCACAACGT 1150
1151  TACACCTAGC TGCTTTATTT GCTGCTTGCA AGTGTAGCTG GTTTGATCAT 1200
1201  GTTGTGTACG AGTTCTAATT TACGTGTGAC CACTAAGCTT TACTGCAATT 1250
1251  ATCTGCGATC TTTAAATGTT ATCCTTTGTT GGAAGGTTAT TATGGTTGTA 1300
1301  TAGCTTCTGT GATATCACGA TTCTGAACAA ATCAAATGTT TGCTGTAGCT 1350
1351  TACAAAGTTT TTGTGATTGC AGAACCACTG TGTTCAGCCA CTCCCAGACT 1400
1401  GTGGTTGTGT GCCCAGGCTG CCAAACTGTT CTCTGCCAAC CTACTGGTGG 1450
1451  GAAGGCCAGG CTCACCGAGG GGTGCTCCTT CCGTCGCAAG GGCGATTAGA 1500
1501  TCCTGCTGTC TTTTAATGGG AGAAGAGAGA AATTTGCTGT TTTACAACCT 1550
1551  TCCCATGATA TGTACTCGTT GAGGATTTTG TTAATTATTA TGACTGTTTA 1600
1601  GCTTGCCCTG TCAATGCTTT TGTAAACGTG CACTTTCCTT GAGCTTAGTG 1650
1651  ACATCTACTT AAGGTGCTGT TTGGTTTTGC CAAGGGATTG GCAGTGGTAG 1700
1701  TGAAATCAGT TGCTGACGTT ATTTGTTGGG GTTTAACTGA AAGTGTCATT 1750
1751  TGTTTAGTTG CCTTGTGTAA TGGATTCAAT TTAACAACTG ACTAGTACGG 1800
1801  TCTCTAGCCC AATGCAGCGT TCACTCTCAG CGGCAGATTG TTCTGTAGCA 1850
1851  GTTAACGGTC ATAGGAGTGC CCTCGCGCTC GCCTTTGTCC CCTCTCTCCC 1900
1901  TTCCTCCTAT GGAGGAACGC CCGAGGCGAC CGCATCTCGA AGATG  1945
```

Fig.20

```
   1  CTGCAGCAAC AGATAATAAC GAAATCAAAC CCTCAACAAA CCTAATAAAA   50
  51  AAAATACTAA ATGGTGCTAT CTAATACCCA GTTCTGGATA TATGGAGTTG  100
 101  TAGTGCGCCT CAGCCCAAAA ACTAATCATA TAACTGCAAC AGCAAGACTT  150
 151  TAGATAGCAA CACCACGAAT CAGTCTTCTA AAACTTCACT AGTTTATTAT  200
 201  CACCAGCGGA TGCAAACAAA GCTCCACTCA CACTGAAGCA AATAGCTCTT  250
 251  ATAGCGTCTG AATGAGAACG ACCACCACAA TCATCTGACA ATGAAACTGG  300
 301  ATGACCAGCC CTGGAGGAAC AGAGTAAATA TTTACTTGCA CATTCCACCA  350
 351  AATGAAATGC GGAGAAAAGG CGCTCCACTT CTCACTGAAA CATGTGAACT  400
 401  CAGGATTCAA CAATATAAAA TTAAGCAACC AGGACCTGAT TTCGTTAGTA  450
 451  ACAATTTACA CAACCTGACT CACAAGAATT CCATTGTTTT CTTAAAGAAA  500
 501  ACTTTCCATT CCTCTCCACA CGAAACTAAG TTATAAGAGA AAACAACTAA  550
 551  CGATAAGCAG CAGATAGATT CAGTTCAGAA ACGCACTACA ACCTCCACCA  600
 601  AAATCTACCA ATCGTGTAAT CAAACACTAG CACTTTCTGC ATAAGAGGTC  650
 651  CATGTTTAAT TAGGAAGCGG TGATTGAAGC GGGTATTCCA ATTTCGAAAC  700
 701  CTGAGATCGA ACACGCGGAG CTCGGGCCCA ATGGCGACGG CGACGGAGTT  750
 751  GCCGTGCGGG TGGGCGCGAC GAGCGCCGGC GCGAACTCCG CCGCGCCGCT  800
 801  TACCCTCGGG CTCTTCGATC GCTGTGTCCT CCATTGCAGC GGCAGCCGCA  850
 851  AGATAGGGAG CTCGGATTCG AATCGAGCGC GAGACGGAGG CGGGGAGGGA  900
 901  TTCAGGGATT AGGGTTTACA GGCAGTCGCA CATGGGGCCC AGAGACCAGT  950
 951  GTCACGAAGG AGAGCCCGGA TGGGTTGCAG TTGCAGGCTT GCAGCCCAAA 1000
1001  GGGCAAAAGC CTTTGGTTTG TCGGTCATGG GCCTCCACGA AAACTGTCTC 1050
1051  TGTCGGACTG CCCAATCCCA CGGAAGGCCG AGATGAACGC AGCACCTCGA 1100
1101  TGAACGCCTC AGATTCGCCA ACCCACACGA CCGCACCTAT ATAAAGTATC 1150
1151  CACCCTCCGC TCCCGTCTTC CTCTCCGCCG CCGCCACGCG CTCGCCTTTG 1200
1201  CTCCCTCTGT TTCCCTTCAT AGAAAGGCGC AGGCCAGGAA CGCCCAAGGC 1250
1251  GACCGCATCT CCAAGATG  1268
```

Fig.21

```
   1  CTGCAGCTAG CTGATGATCG AGACTCCGTT TGAACTTTGA AATGTACGAA   50
  51  TTTTCTTGAA CCATGTGAAA TCGGAGTCAT TTTGCACAAA ACTTGTGAAT  100
 101  GAAGATCCAA ATATTCCTAA CGTATAGAAC TTCAAAGAAA GCTGGAGAAA  150
 151  ATGAAAGAAA GATCACTTAT AGAAAAAAAA AATGAAACTG AACTTTGCCT  200
 201  TGCCACCGAT GACATCAAAA GCATTTTCGC CTAATGCTTG TAGTGGACCG  250
 251  GCCTTGAGCT AGCTGCCATG TGATCGCTGA TCCTTTCGGC AGTGATGAGC  300
 301  TAGCTATGCT ACTTCACTGA AGCGATGATG AGGTGTCTCG CTGCCCCCAT  350
 351  TGCGGTGTAC CACAAAACGA GTCTGACCTC GCTGTCCTGC TGATGGCATC  400
 401  CTTACTTGCT TTTTCCATTA TTTTGCAAGG CAAAGTTGAT CCATGGACAA  450
 451  CTACTCCCGC AGAACAGTTC AATGGGCTCA AATATTTCTA TGCTAGCTTT  500
 501  TTCGATAAAG GTGGTGGTGC CTAATGTTGT CTAAAGCAAG GAGACGGACT  550
 551  TGACCCAAAG TTGATAAGGG TCTCATCCAT TTGCCTTGAT TAAGCGGAAC  600
 601  AAGACACTTG ATAGGAATAG GTTTGGTTTT TACCTAAAAT GATGCACATG  650
 651  AAAGTTATTT TTTGTCAAAC TCCAAATCCT CAAATAGCTT ACCAAAGTTT  700
 701  TGGCCAAATT TAGATTTGAA AATAAAGTAT AGTGTCGAAT AATTATGTTG  750
 751  TCTACCTAAA CTTTTTTTCC ATCAAATAAA AGTTCAGAGT TTTTAGTGGG  800
 801  TGGTGATTGT TATATAGGGG GTCGACACGG AGCTCTTTTA ATGAACTAAT  850
 851  CTAAGTTTTC TAATAATCTA TATCTAATAT CTGTCATCCT TTGTCCCTTA  900
 901  CAACTGTCAG ATGGAGATTT GACGAACTCA ATCCCTTCAA TTCTTATACC  950
 951  CATACAAGCT AGAGCGACAC GCATCTGGGG CACACTGTGG TGTTCGATTT 1000
1001  GCAAACGAAT TGAAACGCAT GATGACATGA TCGCTAAATA AATCTCCAAG 1050
1051  CCGCAAGTCT TCTAGCAAGT AACCACGCAA GAAGTTAATT GTCTTATTGC 1100
1101  AGCGCACGGG TATATTTGCT AGTTATATTA ATAAGAGAAA ATTTCTTCAT 1150
1151  CCAAATTTTT GTTCGTCCCT CTCTCCCGAT CCATGTAACT GTCAACTCCC 1200
1201  TTGAGGGCCA ACAATAAGAG AATAGTGGTA CGTGATGAAT TAAATATAAG 1250
1251  GGTTCTAATA GGTTTACTGT TTTGTGTTGT GTCAAACTCA CCGCCGCGCC 1300
1301  CATATACGAT TCAACTAATC TTTGGATTAC CGTAACTTGA CCTGACTGTA 1350
1351  TCAAAACTCC TTTTATTCTG CTTAATGAAA TACGTGCTAA AGCACGATCT 1400
1401  CGAAAAAAAA CTACAGCACA GTGTCCAATT TCAAGATATA TTAGAGCAAA 1450
```

Fig.21 cont'd

1451	TGATAGAAAT	TGTATATCCT	GTACATTGCC	GCACACGAAA	ATATTTGCTA	1500
1501	ATAATAAAAA	AGATGCCATA	AAATTGCTTG	AAGCTCCTGG	TAATAAGCAG	1550
1551	CTGGTAAATA	ATTCCTTAAA	ACGAGAAAAG	AAGAACCTCA	CTAATATCTC	1600
1601	CTTTGCTTAC	CTTCAGTTCA	TCAGCCAACG	ACGAGGTAGG	GTTCAGTATC	1650
1651	ATGATCCAAT	TATCCCATCG	TGACATAGCC	TTGTCCTTGA	TGATTCGAGA	1700
1701	TGCAATTCTA	ATTCTCATCA	TATCATCGAC	TAGGTAACAC	AGAAAACAAA	1750
1751	CTTTTTTCTT	CGTCAATTGC	ACTGCAACCG	TTGCTTTTTG	TGATGTGCAG	1800
1801	TTGTGCACCC	ACATCACAAC	AGCACTCAAC	ACATGCCACT	TCAAAGTTCG	1850
1851	AATCGACACC	AGAACTGACG	GGAGAAAAGA	AAACAAATTA	ACAAAACTGT	1900
1901	AGAATAGATC	ATCCAGTCAT	CCAGCGTCCA	AAAAGTCCTG	CTAGCTATAG	1950
1951	ATGCAACCTA	ATAACTTTGC	TGACCTAGTC	ACTCCGAATT	CCAACATCAC	2000
2001	ATCATCGTAG	TAGGCTCATT	GTCATAGCAT	TCCTCAACAG	CACTGTTAAC	2050
2051	AAGCAGATGC	AAACAAGCAG	ATGCACATTC	ACCGGTCCCA	ATTGCACTGT	2100
2101	TAACCCGACA	ACCGCCTCTA	TTCCTGTCAC	CTCTAATCAC	GTACGAGCAA	2150
2151	AGCACTAATC	AATCTCTCTC	CCCTCCCCTG	TAACCAAAGC	TGCCGAATCT	2200
2201	CTTCGCTGAT	CTGGCTGCTG	CACCGCTGCA	GCCTGCGAGG	AGAGGTGGGG	2250
2251	GTCGCTCATC	AGTTAACTAC	TGACGCCAAG	GACGGCGTGC	AACGTGCAGC	2300
2301	AACAATGAGG	CGCAGGATGC	ACACCTCACA	ATGCCCTGCA	CCGCAATGAA	2350
2351	GCCTACATCT	CCCGAAGGTA	CACTTGTCTC	CATGGACACA	CATGCTGGAA	2400
2401	CCTGGAGAGA	TGCATGCAAG	CAAGCAGACC	ACACATCATC	GATCGGTCTG	2450
2451	TGAACCTGTG	ATCACTTGGT	GTTGGTGATC	TATCGATCTC	TCCCACAGAT	2500
2501	TCACGCACAG	GGGCCTGCTG	GGTGAAAGAA	CCAAGACACC	GTGCACACCG	2550
2551	CCCCTTTTGG	ACCCTTTCCA	TGTGTCATGA	CACGGTCATG	GGCACCTTTT	2600
2601	GGTTGCATTG	CATGACATGT	TCATGTGTGC	TCGTACAGCT	CCTCAAGATT	2650
2651	CCCGATCATG	ATACCAGCGA	CGACAGCTCT	TTATTAGATC	AAGGGGGATT	2700
2701	TTTAAAAAAA	AAATCGGCAG	AGTGCTAAGA	AGCCCTTTTG	TGTTCTCCCC	2750
2751	ATTCTGGCCT	CGCCGCCGGG	CCTCTATATA	TAGCGCTCCC	ATCTCACCAC	2800
2801	CTTGCTTCAC	ACAAGCTCGA	ATGCTCGATC	GATCCCACTC	TCAGCTGATC	2850
2851	GCTCACTCTT	GCAGCTCGAT	CAGTCTTAGC	TCTAGCCTCT	AGCTAGCCAA	2900
2901	CTAGCCACTC	CTTCGTGTAG	CCATCAGCCT	TCTGATCGTC	ACCAATG	2948

Fig. 22

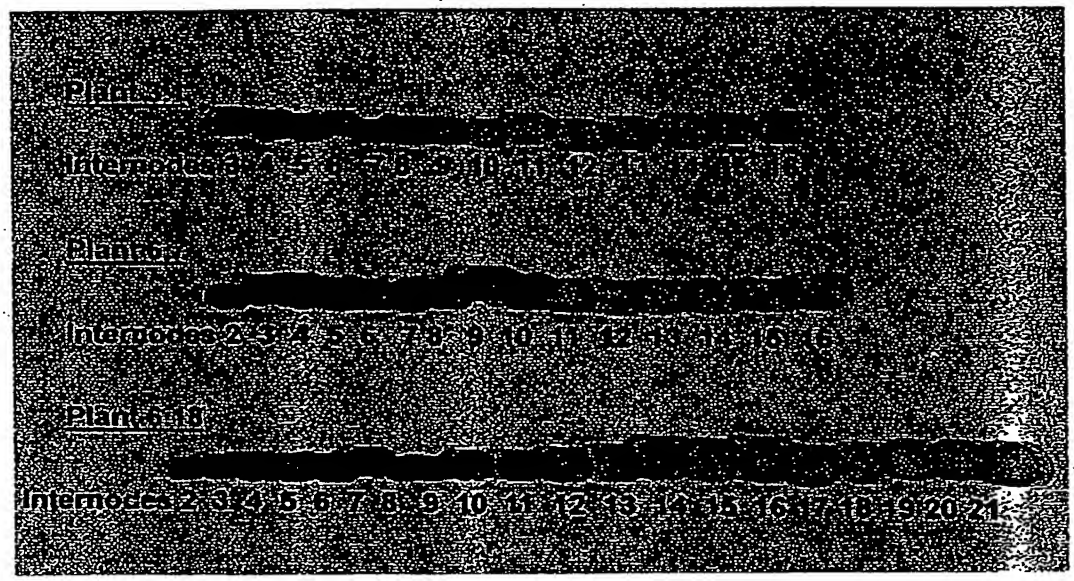

Plant 1
Internodes 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Plant 2
Internodes 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Plant 10
Internodes 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Fig.23

```
  1  TCCTTTATTT CATTCATTCG TTCATCCGTG TGAATTGGTA CAGGTGCATA   50
 51  AAGAAAAGCT TAGCTAGCCC GCCATTTTTA GGCAGCCGGG CAGGTATACG  100
101  GTGTTGACAT ACATACATAC AAGTAGGACG GATTTACAAC GCTAGTTGTG  150
151  GCAGATTTGA ATAAAAAGAA GTAGCAAATT ACACGCGAAT TCGAATCCAT  200
201  TCAGTTGTCG AGGTGTACTA TATAGCAAGC TCATCGGAAG ACGACGACGA  250
251  CGTACGGTTC GTTCGTTCGT TTATTGACAA ATAATTCCAA ACTTGCACCT  300
301  GGATGCATGA TGATATACGT ACACGTGTAC GTACTAGCAG GTTGGATCGG  350
351  ATTGATCACC AGGAGA  366
```

Fig.24.

```
  1  TTGAGTACTG AGTAGCTAGC TAGCCCAACT GTTTACTTGC TTTCCCTTTC   50
 51  GCGAGCTGGG TGGGGTGGTT TGCGCTTCAA AGGCCGCGGC GCTTCTCTGG  100
101  CGGCCTCACG CCGGAGGCCT TGCCGTCGCT GAAGAAGCTT TTGTCGGGGA  150
                                                    S  G  K
151  AGCTGGAGAA CAACCAGATC CCTTGGCGTG GGGACTCAGC GCTGACCGAT  200
       L  E  N  N  Q  I  P  W  R  G  D  S  A  L  T  D
201  GGGAAGGAAG CGGGACTGGA TCTTTCCAAG GGAATGTATG ATGCTGGGGA  250
     G  K  E  A  G  L  D  L  S  K  G  M  Y  D  A  G  D
251  TCATATGAAG TTCACCTTCC CGATGGCATT CACGGCGACG GTGCTCGCGT  300
      H  M  K  F  T  F  P  M  A  F  T  A  T  V  L  A  W
301  GGTCGGTGCT GGAGTATGGG GATCAGATGA GCGCGGCAAA GCAACTGGAC  350
       S  V  L  E  Y  G  D  Q  M  S  A  A  K  Q  L  D
351  CCTGCCCTCG ATGCACTGAA GTGGATTACT GATTCCTTAT CGCTG  395
     P  A  L  D  A  L  K  W  I  T  D  S  L  S  L
```

Fig. 25

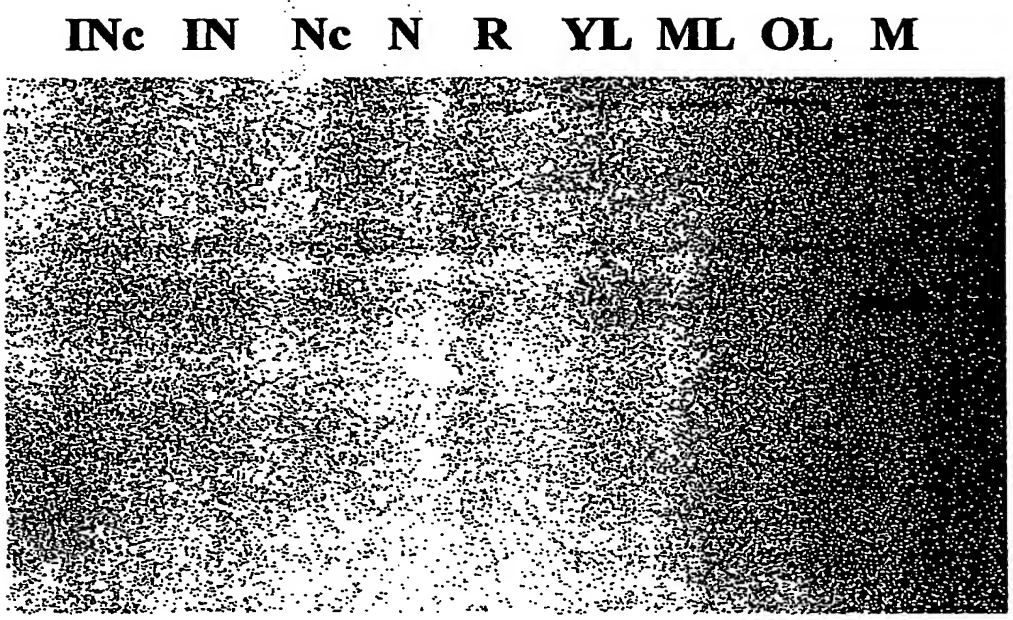


Fig.26

```
1   ATGTCGTGCT GCGGAGGCAA CTGTGGGTGC GGCTCCGGCT GCAAGTGCGG   50
    M  S  C  C   G  G  N  C  G  C   G  S  G  C   K  C  G
51  CAGCGGCTGC GGAGGGTGCA AGATGTACCC GGACATGGCT GAGCAGGTGA  100
     S  G  C   G  G  C  K   M  Y  P   D  M  A   E  Q  V  T
101 CCACCACCAC CCAGACTCTC ATCATGGGTG TTGCACCATC CAAGGGGCAC  150
      T  T  T   Q  T  L   I  M  G  V   A  P  S   K  G  H
151 GCCGAGGGCG GGTTCGAGGC GGCGGCCGCC GGAGCTGAGA ACGACGGCTG  200
    A  E  G  G   F  E  A   A  A  A   G  A  E  N   D  G  C
201 CAAGTGTGGC CCCAACTGCA GCTGCAACCC CTGCACCTGC AAGTGAGCTG  250
     K  C  G   P  N  C  S   C  N  P   C  T  C   K
251 ATGACCCTGC CGTGATGAGG AGATCGGAAG GACTCTAGCT AAGCTCTATT  300
301 ATAGCTGTGT CGTTTGTGTC TCAGTCAGTC AGTCGCGTGC CATGGCATGC  350
351 GTAGTTGCTC GCTCGCATGC CGCTAGTGGT GGTATTGATG GGAACCAATA  400
401 ATTGTACTGA GCCGTCGAGT CCTGTCCTTC CTTCCCCTGC CCCTGGGGCT  450
451 CATCTAGCAG TCGGTGTTGT GTTTGTGGGT GCAGCACGGC AGCAGGCCAT  500
501 GCCCCATGCT CTGTCTGCAC GTCTTGTGCG GCTTTGTTTG TGCTGTGTTA  550
551 CCGTCCATCT ATATCTATAT CCCATCTTTG TGAACAAAAA AAAAAAA  597
```

Fig. 27

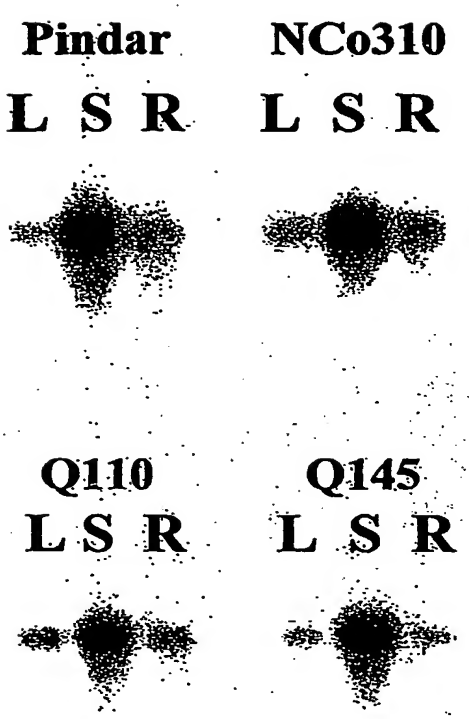


Fig. 28

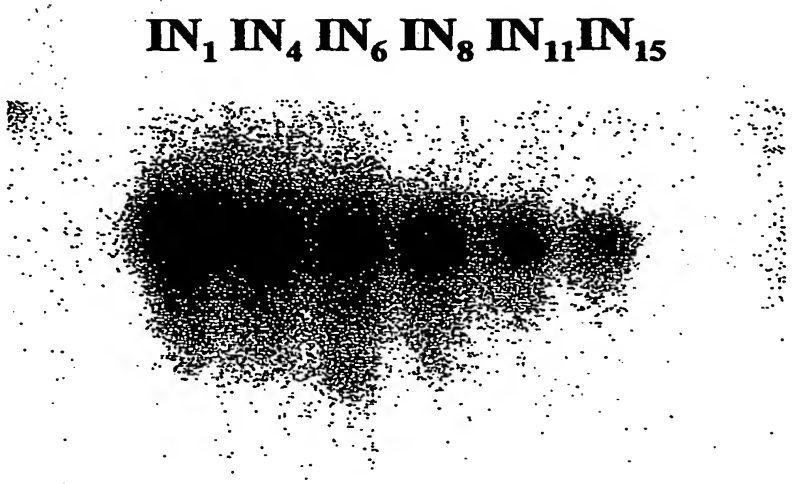


Fig.29

```
   1  CCACACGAGG ATCATTCACT CACTACAGTG GCGTACTAGA ATACCACTAG    50
                                 M  E  D  L  Y  S  I  H  P
  51  TGCGCACACA CGAGCAGCAG CCGCCGTGCA TAGAGTAGTT GTGATGTACA   100
  101 GGTAGTAGCA GCAGCTCGGC TCCATGGAGG ATCTGTACAG CATCCACCCG   150
   G  I  S  R  V  G  G  A  A  S  E  A  S  T  A  G  V
  151 GGGATCTCGC GGGTCGGCGG CGCGGCGAGC GAGGCGTCCA CCGCCGGCGT   200
      G  A  G  G  P  S  P  S  D  L  T  E  L  M  K  A  Q
  201 CGGCGCGGGC GGCCCCTCGC CGTCTGATCT GACGGAGCTC ATGAAGGCGC   250
      I  A  S  H  P  R  Y  P  S  L  L  S  A  Y  I  E
  251 AGATCGCCAG CCACCCTCGC TACCCCTCCC TCCTCTCCGC CTACATCGAG   300
      C  R  K  V  G  A  P  P  Q  V  A  S  L  L  E  E  V
  301 TGCCGCAAGG TGGGAGCGCC TCCGCAGGTG GCGTCGCTAC TGGAGGAGGT   350
      S  R  E  R  S  P  G  A  A  G  A  G  E  I  G  V  D
  351 CAGCCGGGAG AGGAGCCCCG GTGCCGCCGG CGCCGGGGAG ATCGGCGTCG   400
      P  E  L  D  E  F  M  D  S  Y  C  R  V  L  V  R
  401 ATCCCGAGCT CGACGAGTTC ATGGACTCTT ACTGCCGGGT GCTGGTGCGC   450
      Y  K  E  E  L  S  R  P  F  D  E  A  A  S  F  L  S
  451 TACAAGGAGG AGCTGTCGCG GCCGTTCGAC GAGGCCGCGT CGTTCCTGAG   500
      S  I  Q  A  Q  L  S  N  L  C  S  A  G  S  S  P  A
  501 CAGCATCCAG GCGCAGCTCA GCAACCTCTG CAGCGCCGGC AGCTCGCCGG   550
      A  T  A  T  H  S  D  D  M  M  G  S  S  E  D  E
  551 CGGCGACCGC CACGCACTCC GATGACATGA TGGGGTCGTC TGAGGATGAG   600
      Q  C  S  G  D  T  D  V  P  D  I  G  Q  E  H  S  S
  601 CAATGCTCAG GGGACACTGA TGTGCCAGAC ATAGGGCAAG AACATAGCTC   650
      R  L  A  D  H  E  L  K  E  M  L  L  K  K  Y  S  G
  651 TCGCTTAGCT GACCACGAAC TCAAGGAAAT GCTTCTTAAG AAGTACAGTG   700
      C  L  S  R  L  R  S  E  F  L  K  K  R  K  K  G
  701 GATGCCTCAG CCGTCTTCGT TCGGAGTTCC TGAAGAAGCG GAAGAAAGGG   750
      K  L  P  K  D  A  R  T  V  L  L  E  W  W  N  T  H
  751 AAGCTACCGA AGGATGCACG GACAGTATTA CTAGAGTGGT GGAACACGCA   800
      Y  R  W  P  Y  P  T  E  E  D  K  V  R  L  A  A  M
  801 CTACCGCTGG CCTTATCCTA CGGAGGAAGA TAAGGTGAGG CTTGCAGCGA   850
  851 TGACCGGTCT CGACCCAAAG CAGATCAACA ATTGGTTCAT CAATCAGAAG   900
```

Fig.29 cont'd

```
      T  G  L   D  P  K   Q  I  N  N   W  F  I   N  Q  K
 901  AAGAAGCATT GGAAGCCATC CGAAGACATG CGGTTCGCGC TCATGGAGGG  950
      K  K  H  W   K  P  S   E  D  M   R  F  A  L   M  E  G
 951  TGTTGCCGGT GGATCTTCTG GGACGACATC TACTTCGATA CAGGCACAAT 1000
       V  A  G   G  S  S  G   T  T  S   T  S  I   Q  A  Q  L
1001  TGGACCCTGA ATCACACACC ATTTGGGATG ACAATTGGCC AATTCAATCA 1050
      D  P  E   S  H  T  I   W  D  D   N  W  P   I  Q  S
1051  GTAGTAAGAC CTGGCATGTG AAGTGACGAT CTGCCCGGTC AAAATTGACA 1100
      V  V  R  P   G  M
1101  ATAAATCTGT CGAGCTGAGG TTGATCACAT TAGTCAGTTG CCCCAGATCA 1150
1151  TGTGTATATG GTGCCATCGT ATCAAAACAA ACTGTATGTA TGGGCGAATT 1200
1201  GAGGAGACCT GCAAAAGCAT TTAATTAGTA GTTTCACGTA TTGGCTCATG 1250
1251  GATTTGTAAT ACTCGCTACC CAATTTAATT TTTAGATAGG CTGAAGGGCT 1300
1301  TATAAAGATA AAATTACTTC TGCAAAAAAA AAAAAAAAA 1339
```

Fig. 30

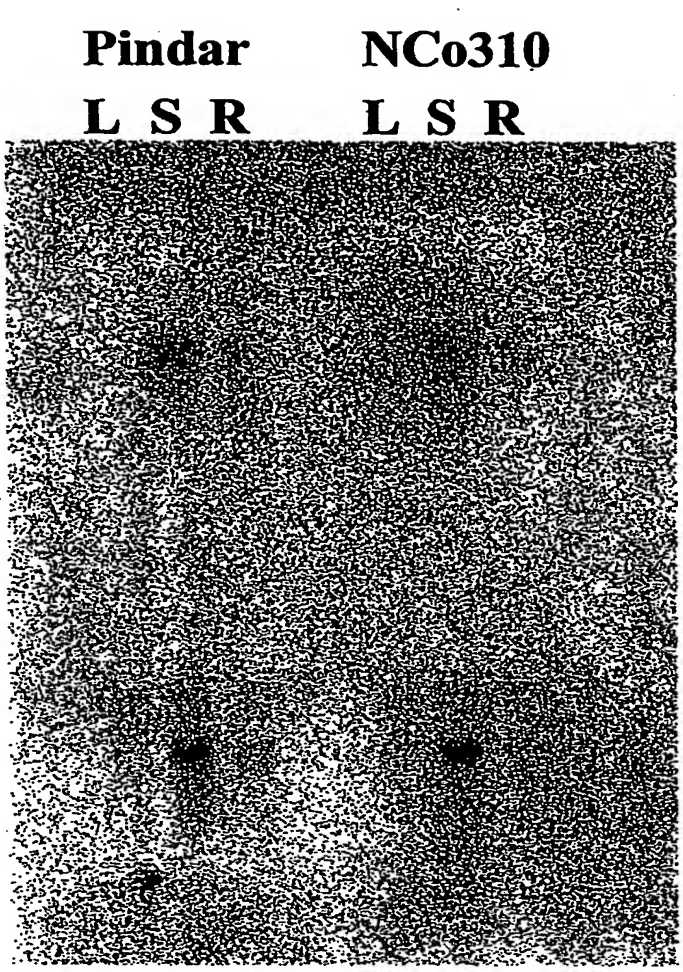


Fig.31

```
   1  TGAGCAGCGT CGTCACCGCC GCCGCCGCCA TTCAGGCTGA CCCCGCCGAG   50
        S  S  V  V  T  A  A  A  A  I  Q  A  D  P  A  E
  51  AAGCCGCGCG CCCCCAAAGC TCAGGTGGTG GGATGGCCAC CGGTCCGGTC  100
      K  P  R  A  P  K  A  Q  V  V  G  W  P  P  V  R  S
 101  GTTCCGGAAG AACATCATGT CGGTGCAGTC CGACAAGGGC GCCGGCGGCA  150
        F  R  K  N  I  M  S  V  Q  S  D  K  G  A  G  G  S
 151  GCAAAGACGC CGACAAGTCC TCGCCGCCGC CGGCGGCAGC AGCAGCAGTT  200
        K  D  A  D  K  S  S  P  P  P  A  A  A  A  A  V
 201  GGCGGTGCCG CGTTCGTGAA GGTGAGCTTA GACGGCGCGC CATACCTGCG  250
      G  G  A  A  F  V  K  V  S  L  D  G  A  P  Y  L  R
 251  CAAGGTGGAC CTCAAGATGT ACAAGATGTA CAAGAGCTAC CAGGAGCTGT  300
        K  V  D  L  K  M  Y  K  M  Y  K  S  Y  Q  E  L  S
 301  CCAAGGCGCT CGAGAAGATG TTCAGCTCCA CCATTGGAAG CTGTGGGTCT  350
        K  A  L  E  K  M  F  S  S  T  I  G  S  C  G  S
 351  CAAGGGATGA ACGGCATGAA CGAGAGCAAG CTGGTGGATC TGCTCAACGG  400
      Q  G  M  N  G  M  N  E  S  K  L  V  D  L  L  N  G
 401  CTCCGAGTAC GTGCCGACCT ACGAGGACAA GGAGGGCGAC TGGATGCTCG  450
        S  E  Y  V  P  T  Y  E  D  K  E  G  D  W  M  L  V
 451  TCGGCGACGT GCCGTGGGAG ATGTTCGTCG AATCATGCAA GCGCCTTCGG  500
        G  D  V  P  W  E  M  F  V  E  S  C  K  R  L  R
 501  ATCATGAAAG GATCAGAAGC CATTGGCCTC GCGCCGAGGG CCATGGAGAA  550
      I  M  K  G  S  E  A  I  G  L  A  P  R  A  M  E  K
 551  ATGCAAGAAC AGAAGCTGAG GAGAAGATGG AAGCTTGCAT CTGAACACCA  600
        C  K  N  R  S
 601  AGGATGGATG CAAGAAGATC CATCTCCACA GCTTTTGCAG CTCAGCTCAT  650
 651  CTCTGTATCA TACTCGTCTC TGTCCTCCAA GCCTCTATAT CTCCTGATCA  700
 701  ATGAAGTCAT TTTGGTGTGT CAACTGTTCT ATCTGTGCTA GTTTGTCAGT  750
 751  AAACGAGTCA GTCAGTGTTC AAGTCTCATC TTGTCCTATA CAGAAGAGAG  800
 801  AAGACATGAG CATGTGTACC TGTCCTATAA CAATGTCTTT TTGCACTACC  850
 851  AGTTCTCTCC TCTTCTTCTT TTTTTCTTTT TTTTTGGTGG CACAAAATGC  900
 901  AAAGGCACCC TGCTCCTAAT CAAGGGGCCC TTTTGCATCT TGTCCATCAA  950
 951  AAGTTGTTTA CTTATAATTA TCACGGTGTG TCAGTGGAAC AAATAATTGT 1000
1001  TGCAAGTTGT TAAAAAAAAG CTGTATTGTA TGAAAAAAAA AAAAAA 1046
```

Fig. 32

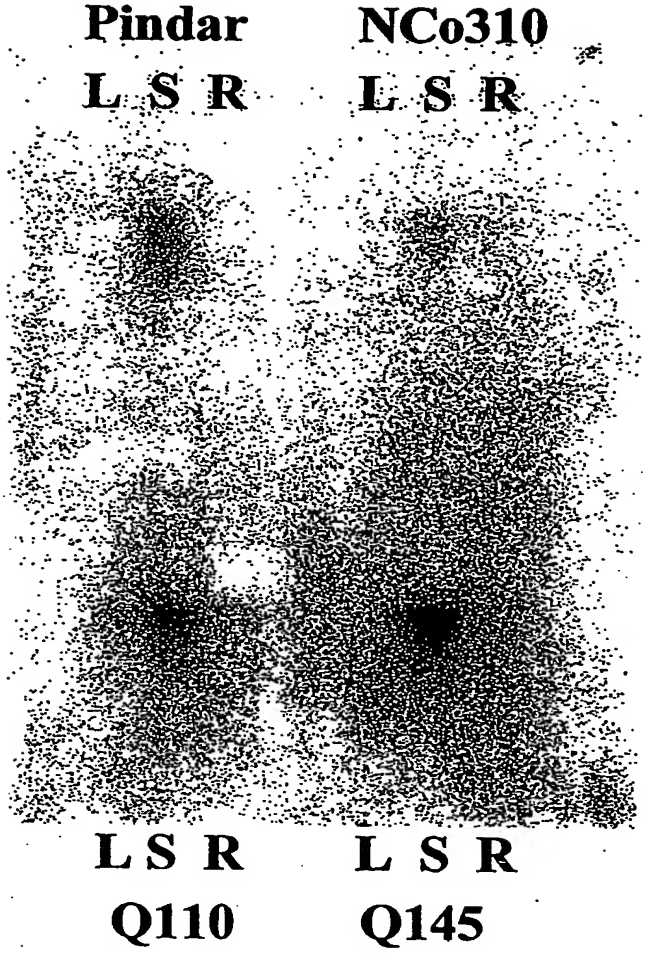
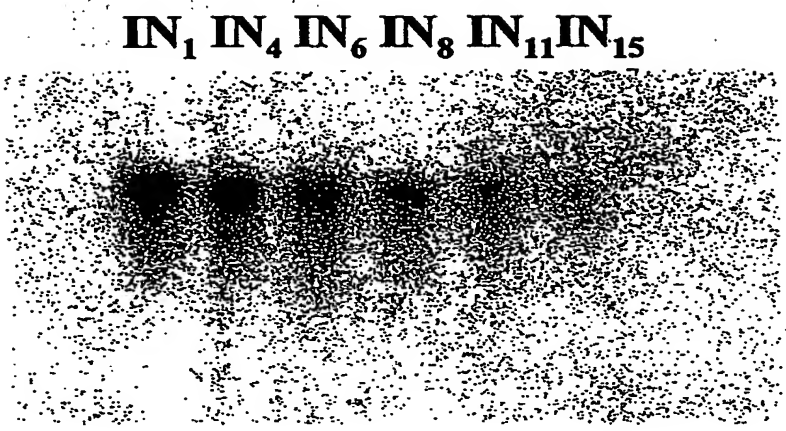


Fig. 33



INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU00/01033

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: C12N 15/29, C07K 14/415, A01H 5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
AS BELOW.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AS BELOW

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EMBL GenBank : Seq ID.Nos. 7, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
DGENE : Seq ID Nos. 16, 17, 18, 19, 20, 21, 22, 23

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -	EMBL, Accession No. AF111029, A novel maize gene homologous to 40S ribosomal protein S27., Stukart G. C. *et. al.* 21 December 1998	38,39,40
X -	EMBL, Accession No. AF083336, Involvement of *Arabidopsis thaliana* ribosomal S27 in degradation triggered by genotoxic stress., Revenkova E., *et. al.* 29 January 1999	38,39,40
X -	GenBank, Accession No. HVMNE1MR, Zinc-finger protein., Huang, C. *et. al.* 16 March 1995	38,39,40

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 October 2000	-6 NOV 2000
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer ALISTAIR BESTOW Telephone No : (02) 6283 2450

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU00/01033

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL, Accession No. AC006434, Genomic sequence for *Arabidopsis thaliana* BAC F10A5, complete sequence., Cheuk R., *et. al.*, 25 January 1999	38,39,40
X	GenBank, Accession No. AC007797, *Arabidopsis thaliana* chromosome I BAC F6F9 genomic sequence, complete sequence., Nguyen M., *et. al.* 15 June 1999	38,39,40
X	DGENE Accession number 1999P-Y11496 and AU, A, 85554/98 (GENSET) 11 February 1999. See specification and sequence listing.	38,39
X	AU, B, 13087/97 (725340) (GIST-BROCADES B.V.) 3 July 1997 See SEQ. ID No. 28.	38,39
X	US, A, 5 243 041 (JOSE A. FERNANDEZ-POL) 7 September 1993 See SEQ ID Nos. 1 and 2.	38,39

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU00/01033

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
AU	85554/98	EP	1000149	WO	9906551		
AU	13087/97	WO	9723633	EP	870042	CA	2241267
		EP	780474				
US	5243041	NONE					

END OF ANNEX